AN EXPERT SYSTEM FOR FREQUENCY ANALYSIS

By

RAMA CHANDRA RAO KALAGA

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

December, 1991

AN EXPERT SYSTEM FOR FREQUENCY ANALYSIS

A Thesis Submitted
In Partial Fulfilment of the Requirements
for the Degree of
MASTER OF TECHNOLOGY

By

RAMA CHANDRA RAO KALAGA

to the

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

December, 1991

to

my

beloved parents

and

grandmother

6/1/91

C E R T I F I C A T E

This is to certify that the thesis titled " **AN EXPERT SYSTEM FOR FREQUENCY ANALYSIS** " submitted by Shri Rama Chandra Rao Kalaga, in partial fulfilment of the requirements for the degree of Master of Technology of the Indian Institute of Technology, Kanpur, is a bonafide research work carried out by him under my supervision and guidance The work embodied in this thesis has not been submitted elsewhere for a degree

December, 1991

(S Ramaseshan)

Professor

Department of Civil Engineering

Indian Institute of Technology

Kanpur - 208016

ACKNOWLEDGEMENTS

I express my deep sense of gratitude, gratefulness, respect and admiration to my thesis supervisor Dr S Ramaseshan, for the successful completion of my thesis work I consider it my privilage to have been his student

I thank all my professors in IIT Kanpur for being friendly and cooperative

I thank Jaykumar for helping me in my course work as well as in my thesis work and for giving me a lot of friendly advice and cooperation I thank Ravi Sankar for helping me in every way and for being a good friend I also thank Reddy, Purwar and Ramesh Kumar for helping me in my thesis work

I thank my fun loving friends Hari, Venu, Konda, Murali, Ramesh(Durga, Totti, Maruti), Arroit, Subba Rao(Kalari, Manne), Kantham, Sharma, Satty,Rao(PT,TS), Balram, Das, Chaitanya, Prasad(YJJ,A), Ravi Chander, Choube, Vivekanand, Mohapatra, Rajeev, G.P Sing, Sailesh, Anupama, Sadhana, Prabha and Nazima for making my stay here a pleasant experience

I sincerely thank my Parents and Grandmother for giving a lot of encouragement and moral support in all my endeavours I thank my brother Chekki who has really taken interest in my carreer and gave a moral boost in my studies

I thank the families of Sri Y K Rao and Sri N Swami for hosting great dinners and for making me feel very much at home I thank Divya and Arunasree for being good friends I thank Hima for all her troubles in preparing good dinners and for being a real good friend I thank Ashu for being a great friend and a lovely company to be with and for making my stay here memorable with her good sense of humour.

December, 1991 Rama Chandra Rao Kalaga

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

χ^2	Chi-square statistic
C_k	Coefficient of kurtosis
C_s	Coefficient of skewness
CV	Coefficient of variation
ε	Displacement parameter for Pearson Type 3 distribution
P	Exceedence probability
K_n	Frequency factor for Normal distribution
K_T	Frequency factor for Pearson Type 3 distribution
N	Length of data
u	Location parameter for Extreme value Type 1 distribution
μ	Population mean
σ	Population standard deviation
σ^2	Population variance
$\bar{x}$	Sample mean
s_x	Sample standard deviation
s_x^2	Sample variance
α	Scale parameter for Extreme value Type 1 distribution
λ	Scale parameter for Pearson Type 3 distribution
β	Shape parameter for Pearson Type 3 distribution
Z	Standard normal variate

ABSTRACT

Frequency analysis of hydrologic data is a knowledge domain where lot of decisions have to be taken based on human expertise, intuition and subjective judgement Frequency analysis being data dependent is subject to many errors Moreover, there is no fixed distribution that can be fitted to a given set of data. An earlier study dealt with the structure and development of a ES based computer program FACHVES The Expert system is embedded in the FORTRAN program The program is written in FORTRAN language and the expert system has been developed for the VAX-VMS environment This study deals with modification and enhancement of capabilities of the earlier version

In the present version, a number of subroutines have been added which helps in a more detailed analysis of the data set For example, the subroutine OUTLIER identifies the number and the values of the outliers present in the data, the subroutine CONFBND is a plotting subroutine which plots the fitted distribution, the sample data and the confidence bands. Other subroutines include FREQPL and TPLOT for plotting, MLS for method of least squares and GOFITST for comparing theoretical and actual Chi square values in the goodness of fit test In the given program

i. a distribution can be fitted to the data after taking Expert system advice,

ii the given set of data can be tested for outliers,

iii if outliers are present, parameters can be re estimated by MLS

iv. the goodness of fit tested by Chi square test and

v the fitted distributions visualised graphically by seeing the plots and

vi which distribution is more appropriate to the given set of data may then be decided

Hence, the modifications done has enhanced the capabilities in fitting a distribution

The developed package of FACHVES Ver 2 Ø has been tested with three sets of data representing recorded data at two different sites in India The results of this study verify the

satisfactory performance of the package ,in particular the enhancements of outlier test and graphic output

CHAPTER 1

INTRODUCTION

1.1 General

In this age of computers , engineering design and analysis attained greater heights Fast and efficient computing made several complex analysis techniques , e g , numerical analysis, statistical analysis etc , easy The programs for engineering problems written in higher level languages like FORTRAN, BASIC, C and PASCAL , were able to solve complex problems involving a number of ordinary and partial differential equations, logical constraints , or analyse problems involving uncertainties Development of languages like LISP and PROLOG, in which logical approaches were easier to be programmed , facilitated the design of a system which could tackle the problems that depended upon intuition and subjective judgement of the analyst and could not be solved by routine analytic techniques This led to the development of domain specific systems after the realisation that creating universally wise systems was not possible Subsequently, it was realised, that while expertise may be different from problem area to problem area, the structure of the problem may be the same and so several Expert System shells were developed It may generally be possible to develop and implement Expert Systems in a problem area according to the grammer of the shell used

Water resources engineering deals with many processes which vary in space, time and also probabilistically Many a time decisions are to be taken on the basis of past experience, intuition etc Moreover data available for analysis are also subject to many observational or computational errors So problem solving techniques in this field of engineering design also need such a system which can incorporate the domain specific expertise with computational techniques

Thus, in computer aided analysis and design in engineering, numerical, analytical and logical capabilities of a digital computer are extended with the heuristic capabilities of artificial intelligence and expert systems in efficiently solving problems in the areas of analysis and design

1.2 ES in Water Resources Engineering

Artificial Intelligence (AI) is that part of computer science concerned with designing intelligent computer systems, i e , systems that exhibit the intelligent characteristics of human beings like understanding language, learning, reasoning, solving problems and so on AI is a way to impart intelligence to a computer (Levine, 1988)

The field of AI encompasses robotics, game playing, the automated translation of language and Expert System (ES) ES is concerned with the development of computer software that can partially represent human knowledge and utilise that knowledge to solve complex problems within a specific domain (Johnston, 1985)

Water resources engineering deals with planned development and management of water, the spatially and stochastically varying natural resource above and below ground A large number of uncertainties which can arise due to hydrology, hierarchical and multistage nature of decision making, financial and economic variabilities and implementation techniques, social constraints and changing national conditions, are involved in designing a water resource system, viz , planning, development, design and management So, a large amount of domain dependent expertise is used in planning, design, construction and integrated operation of water resource systems and they are often heuristic in nature

ES are being developed in areas of water resources like database management, information systems and water quality management (Smoller, 1985 , Datta and Peralta, 1986 , Arnold et al, 1989 , Simonovic, 1989 , Datta et al, 1990) and selection of design data as storm (Nielson, 1986) or flood estimation (Fayegh and Russel, 1986) and appropriate treatment technology for water supply / sewage (Arnold, 1986) ES has also been applied for data analysis (Wilson, 1986) , hydrologic modelling and parameter estimation (Engman et al, 1986 , Delluer, 1988) , tank irrigation system (Oswald, 1989),and choice of model to be used and preparation of input data for a reservoir system (Lindberg and Nielson, 1986) Also management of multipurpose system of reservoirs integrated with conjunctive use of surface water and ground water are complex problems which need human expertise,

common sense and heuristics ES like SERPES for sewage rehabilitations planning process and WADNES for handling emergences in a water supply network (Ahmad et al, 1989) have been successfully implemented

The concept of combining the decision making capabilities of an expert decision support system and a Geographical Information System (Intelligence GIS) was proposed by Arnold et al (1989) The primary purpose of such a system is to incorporate heuristic knowledge regarding measurement and other uncertainties in the estimation of hydrologic or water quality variables These applications demonstrate the possibilities of ES applications to water resource engineering

1.3 Objectives of the Study

Frequency analysis of hydrologic data is a preliminary step in the design of hydraulic structures, control of extreme hydrologic events, management of water resources etc This field is suitable for ES application as it deals with data lacking in quality and quantity, having observational and computational errors, and sparseness Moreover there is no universally acceptable methodology for fitting a probability distribution to a given set of data and, if at all, only a limited knowledge about the parent distribution of the samples may be available. Physical processes resulting in a high or low precipitation or stream flow are also not well understood Often the presence of outliers, measurement errors or occurrence of rare combinations of physical processes or multiple distributions makes the problem more complicated Hence an efficient approach to frequency analysis requires incorporation of statistical inference tools as well as heuristic knowledge of knowledgeable human experts in this field. A frame work has been developed for frequency analysis of hydrologic data (FACHVES Ver 2 Ø) and is available for the study The major objectives of the study are as follows .

1 to include graphic capabilities and thus enhance the process of fitting an appropriate distribution by visual observation of the data relationships ,

2 to test for outliers in different distributions and its analysis, and

3 incorporate them in an Expert System available for the study (FACHVES)

1.4 Scope of Study

A variety of computer programs in FORTRAN language are available in IIT Kanpur for fitting one or more probability distributions A number of computing environments like the HP 9000 series with a UNIX operating system, a MICROVAX II computer system with a VMS-VAX operating system and IBM PC 's are available for the study The scope of study was limited to

1 enhancement of FACHVES, a Fortran based program with ES tool (CLIPS, as identified in Sec 3 1 4), in a micro VAX-VMS environment ,

2. Graphics facilities on the text screen itself and

3 a limited number of outlier tests

1.5 Organisation of the Study

The study is reported in the following sequence

1 Introduction to Expert systems , applications in water resources engineering , objectives ; scope and organization of the study (Chapter 1)

2 An outline of frequency analysis, e g , about different probability distributions, identification of distributions, estimation of parameters, identification and tests for outliers for different distributions, confidence bands, and tests for goodness of fit (Chapter 2)

3. A brief introduction to expert systems, building expert systems, CLIPS as an ES shell and its application and the salient features of FACHVES (Frequency analysis for continuous hydrologic variables with embedded expert system) which has already been developed (Chapter 3)

4 The modifications to the FACHVES program, structure and introduction to the use of FACHVES Ver 2 Ø (Chapter 4)

5 Data used, the results, discussions and conclusions for FACHVES Ver 2 Ø (Chapter 5) and

6 Summary, conclusions and suggestions (Chapter 6)

CHAPTER 2

FREQUENCY ANALYSIS

2.1 Introduction

Hydrologic variables like evaporation, precipitation, streamflow, etc vary seasonally and also in a probabilistic manner The random variation of the variables can be represented in terms of the frequency distribution of the variables The fitting of a probability distribution to a random characteristic or variable based on available limited data concerning the variable is an important area of study in Hydrology Hydrologic systems are also affected by extremes of random variables such as severe storms, floods, and droughts The magnitude of a random event is inversely related to its frequency of occurrence, very severe events occurring less frequently than more moderate events The objective of frequency analysis of hydrologic data is to relate the magnitude of random variables to their frequency of occurrence through the use of probability distributions The hydrologic data analysed are generally assumed to be for an independent and identically distributed variable and the hydrologic system producing them (e g, a storm rainfall system) is considered to be stochastic, space-independent and time-independent The hydrologic data employed should be carefully selected so that the assumptions of independence and identical distribution are satisfied

The results of frequency analysis can be used for many engineering purposes; viz , for the design of dams, bridges, culverts, and flood control structures, to determine the economic value of conservation and flood control projects; and to delineate flood plains and determine the effect of encroachments on the flood plain

2.2 Probability functions

The *Relative frequency function* $f_s(x)$ may be defined as the ratio of the number of observations n_i in interval i (the feasible range of the random variable is divided into discrete intervals), to the total number of observations n , which is

calculated from sample data

$$f_r(x_i) = n_i/n \qquad (2.1)$$

The sum of the values of the relative frequencies upto a given point is the *cumulative frequency function* $F_r(x)$, which is also calculated from the sample data

$$F_s(x_i) = \sum_{j=1}^{i} f_r(x_j) \qquad (2.2)$$

The corresponding functions for the population are approached as limits $n \rightarrow \infty$ and $\Delta x \rightarrow 0$ (Δx is the class interval) In the limit, the relative frequency function divided by the interval length Δx becomes the *Probability density function* $f(x)$.

$$f(x) = \lim_{\substack{n \rightarrow \infty \\ \Delta x \rightarrow 0}} [f_s(x + \Delta x) - f_s(x)]/\Delta x \qquad (2.3)$$

and the cumulative frequency function becomes the *Probability distribution function*, $F(x)$

$$F(x) = \lim_{\substack{n \rightarrow \infty \\ \Delta x \rightarrow 0}} F_s(x) \qquad (2.4)$$

Since the sample data are available the distributions and the density functions are expressed in terms of parameters of the sample which are assumed to be equal to the population values

2.3 Parameters of a distribution

The functional form of the distribution is expressed in terms of the parameters of the distribution The statistics of any distribution are the Mean, Median, Mode, the standard deviation , range, coefficient of variation, skewness and kurtosis The statistics are determined as follows

Arithmetic mean of a set of observations is their sum divided by the number of observations, e g , the arithmetic mean $\bar{x}$ of n observations $X_1, X_2, \ldots, X_n$ is given by

$$\bar{x} = (X_1 + X_2 + \ldots X_n)/n = 1/n \sum_{i=1}^{n} X_i \qquad (2.5)$$

Median of a distribution is the value of the variable which divides the distribution into two equal parts such that the number of observations above it is equal to the number of observations below it Thus median is a positional average In case of ungrouped data, if the number of observations is odd then median is the middle value after the values have been arranged in ascending or descending order of magnitude. In case of even number

of observations, there are two middle terms and the median is obtained by taking the arithmetic mean of the middle terms

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely So mode is the value of x corresponding to maximum frequency

The standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean For the frequency distribution whose sample values are X_i with a frequency of f_i, for i = 1,2, n, the sample standard deviation is

$$s = \sum_i f_i (x_i - \bar{x}) / \sqrt{(N-1)} \qquad (2\ 6)$$

where $\bar{x}$ is the arithmetic mean of the distribution and $\sum_i f_i = N$.

The square of the standard deviation is called the variance

Range is the difference between the two extreme observations of the distribution Range is the simplest but a crude measure of dispersion of the values of the variable

Coefficient of variation (CV)is the percentage variation in the mean, standard deviation being considered as the variation parameter with respect to the mean

$$CV\ \% = 100\ \mu\ /\ \sigma \qquad (2\ 7)$$

The series with greater CV is said to be more variable than the other. The series having lesser CV is said to be more consistent than the other

Skewness is studied to have an idea about the shape of the curve that can be drawn with the help of the given data. A distribution is said to skewed if

i) The mean, median and mode fall at different points or

ii) The pdf curve drawn for a given set of data is not symmetrical but stretched more to one side than to the other A sample estimate of coefficient of skewness is

$$G_1 = \frac{n \sum_i (x_i - \bar{x})^3}{(n-1)(n-2)s^3} \qquad (2.8)$$

The skewness is positive if the larger tail of the distribution lies towards the higher values of the variate (the right) and is negative in the contrary case (Fig 2 1)

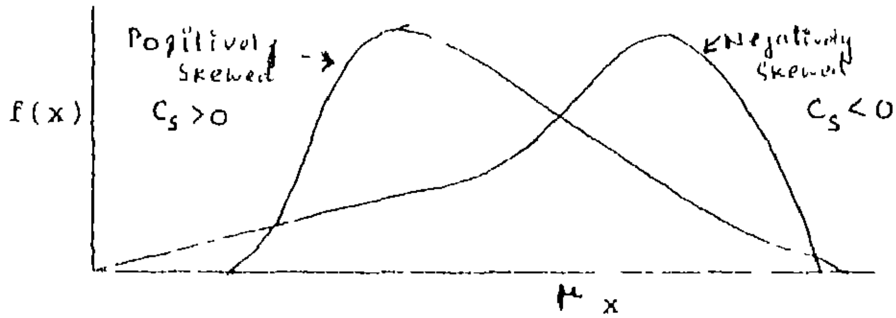


Fig 2 1 Coefficient of skewness

Kurtosis gives an idea about the convexity of the curve. It indicates about the flatness or peakedness of the curve

2.4 Probability distributions

Some of the distributions commonly used in hydrology and their density functions are discussed below

2.4.1 Normal and log Normal distributions

A random variable X is said to have a Normal distribution with parameters μ (mean) and σ^2 (variance) (Sec. 2.4) if its density function is given by

$$f(x\ ,\ \mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\ \mathrm{Exp}\left[\ -1/2\ \left\{\ \frac{x-\mu}{\sigma}\ \right\}^2\ \right]\ , \qquad (2\ 9)$$

$$-\infty < x < \infty\ ,\quad -\infty < \mu < \infty\ ,\quad \sigma > 0$$

The Normal probability curve is bell shaped (Fig. 2 2) and symmetrical about the line $x = \mu$ For large values of σ , the curve tends to flatten out and for small values of σ, it has a sharp peak The mean, median and mode of the distribution coincide The standard Normal variate is given by

$$Z = (X - \mu)/\sigma \qquad (2.10)$$

with $E(Z) = 0$ and $\mathrm{var}(Z) = 1$.

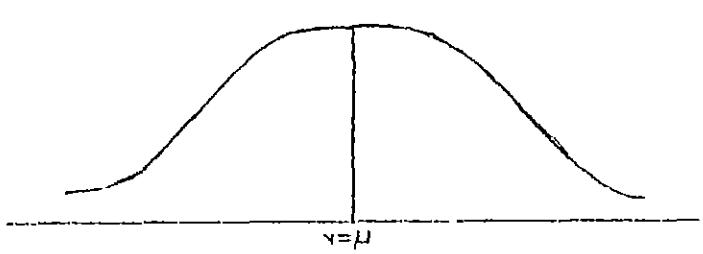


Fig 2 2 Normal probability curve

Hydrologic variables such as annual precipitation are often the sum of many independent events and they tend to follow the normal distribution The main limitations of normal distribution for describing hydrologic variables are that it varies over a continuous range [$-\infty$, ∞] while most hydrologic variables are non-negative and and they tend to be skewed

Consider the exceedence probability, viz , the probability $P(X \geq x)$

$$P(X \geq x) = 1 - P(X \leq x)$$
$$= 1 - F(x) \qquad (2\ 11)$$

Then the recurrence interval, T, which is the average interval between successive equalling or exceedence of x is given by

$$T = 1/P$$
$$= 1/(1 - F(x)) \qquad (2\ 12)$$

The value of Z corresponding to an exceedence probability of P ($P = 1/T$) can be calculated for the Normal distribution by finding the value of an intermediate variable w ,

$$w = [\ \ln (1/p^2)\ ]^{1/2} \qquad (\ 0 < P < 0\ 5) \qquad (2\ 13)$$

and then calculating Z using the approximation

$$Z = w - \frac{2\ 515517 + 0\ 802853w + 0.010328w^2}{1 + 1\ 432788w + 0.189269w^2 + 0\ 001308w^3} \qquad (2\ 14)$$

When $P > 0\ 5$, $1-p$ is substituted for p in the above equation and then the value of Z computed is given a negative sign (Abramowitz and Stegun, 1965)

If the random variable $Y = \log X$ is normally distributed then X is said to be lognormally distributed Chow (1954) reasoned that this distribution is applicable to hydrologic variables formed as the products of other variables since if $X = X_1 X_2 X_3 \ldots X_n$, then

$$Y = \log X = \sum_{i=1,n} \log X_i = \sum_{i=1,n} Y_i \qquad (2.15)$$

which tends to the normal distribution for large n provided that the X_i are independent and identically distributed. The log-normal distribution has been found to describe the distribution of hydraulic conductivity in a porous medium, the distribution of rain drop sizes in a storm, etc The log normal distribution is bounded ($x > 0$) and the logarithmic transformation tends to reduce

the positive skewness commonly found in hydrologic data Some limitations are that it has only two parameters and it requires the logarithms of the data to be symmetric about their mean

Let μ and σ be the mean and standard deviation of the log normally distributed variable The mean $\mu(n)$ and the standard deviation $\sigma(n)$ of the normalised variable are related to μ and σ by

$$\mu(n) = 1/2 \ln \{\mu^4 / (\mu^2 + \sigma^2)\} \quad \text{and} \tag{2 16}$$

$$\sigma(n) = [\ln \{(\mu^2 + \sigma^2)/\mu^2\}]^{1/2} \tag{2 17}$$

2.4.2 Pearson and Log Pearson distributions

Pearson Type 3 distribution, also called the Three parameter Gamma distribution, introduces a third parameter, the lower bound ε, so that By the method of moments, the three sample moments (the mean, the standard deviation and the coefficient of skewness) can be transformed into the three parameters λ, β and ε of the probability distribution

The Pearson system of distribution include seven types, they are all solutions for f(x) in an equation of the form

$$\frac{d\,[f(x)]}{dx} = \frac{F(x)(x - d_0)}{C_0 + C_1 x + C_2 x^2} \tag{2 18}$$

where d_0 is the mode of the distribution (the value of x for which f(x) is a maximum) and C_0, C_1 and C_2 are coefficients to be determined

Pearson Type 3 distribution, also called the Three Parameter Gamma distribution has three parameters When $C_2 = 0$, the solution of Eq 2 18 is a Pearson Type 3 distribution, which has a probability density function given by

$$f(x) = \frac{\lambda^{\beta} (x-\varepsilon)^{\beta-1} e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)} \quad , \; x \geq \varepsilon \; , \tag{2.19}$$

where C_s = coefficient of skewness ,

σ_x = standard deviation of x

$\beta = (2 / C_s)^2$,

$\varepsilon = \bar{x} - \sigma_x \sqrt{\beta}$,

and $\lambda = \sigma_x / \sqrt{\beta}$

The third parameter is the lower bound ε. For Cs = Cs = Ø, the Normal distribution is the solution of Eq. 2.18. Thus Normal distribution is a special case of the Pearson Type 3 distribution, describing a nonskewed variable. The Pearson Type 3 distribution was first applied in hydrology by Foster (1924) to describe the probability distribution of the annual maximum flood peaks.

When the data are very positively skewed, a log transformation is used to reduce the skewness. If log X follows a Pearson Type 3 distribution, then X is said to follow a Log Pearson Type 3 distribution. As a special case, when log X is symmetric about its mean, the Log Pearson Type 3 distribution reduces to the Log Normal distribution.

The location of the bound ε in the Log Pearson Type 3 distribution depends on the skewness of the data. If the data are positively skewed, then log X ≥ ε and ε is a lower bound, while if the data are negatively skewed, log X ≤ ε and ε is an upper bound. The log transformation reduces the skewness of the transformed data and may produce transformed data which are negatively skewed from original data which are positively skewed. In that case, the application of Log Pearson Type 3 distribution would impose an artificial upper bound on the data.

The frequency factor for Pearson Type 3 distribution depends on the return period T and the coefficient of skewness Cs. When Cs = Ø, the frequency factor is equal to the standard normal variate Z. When Cs ≠ Ø, K_T is approximated by Kite (1977) as

$$K_T = Z + (Z^2-1)k + 1/3(Z^3-6Z)k^2 - (Z^2-1)k^3 + Zk^4 + 1/3k^5 \quad (2.2Ø)$$

where k = Cs/6

2.4.3 Extreme value distributions

Extreme values are selected maximum or minimum values of sets of data, for example, the annual maximum discharge at a given location is the largest recorded discharge value during a year and the annual maximum discharge value for each year of the historical record make up a set of extreme values that can be analysed statistically. When the number of selected extreme values is large distributions of the extreme values selected from sets of samples

of any probability distribution have been shown by Fisher and Trippet (1928) to converge to one of three forms of Extreme value distributions, called Types I, II and III respectively

The three limiting forms were shown by Jenkinson (1955) to be special cases of a single distribution called the General Extreme Value (GEV) distribution The probability distribution function for the GEV is

$$F(x) = \exp\left[-\left(1-k\,\frac{x-\mu}{\alpha}\right)^{1/k}\right] \qquad (2.21)$$

where, k, μ, α are parameters to be determined

For k = Ø, the distribution is the Extreme value Type 1 distribution whose probability distribution function is

$$f(x) = \frac{1}{\alpha}\exp\left[-\frac{x-u}{\alpha} - \exp\left(-\frac{x-u}{\alpha}\right)\right] \qquad (2.22)$$

$$-\alpha < x < \alpha ,$$

where $\alpha = \sigma_x \sqrt{6}/\pi$ and $u = \mu - 0.5772\,\alpha$

For Extreme value Type 1 distribution, x is unbounded For k > Ø, the distribution is the Extreme value Type 3 distribution for which the GEV equation applies for $-\alpha \le x \le (u + \alpha/k)$ In both the cases, α is assumed positive For Extreme value type 3 distribution, x is bounded from above (by $u + \alpha/k$)

2.5 Estimation of parameters

The parameters may be estimated generally by one of the three following methods (Chow, 1964 , Yevjevich, 1972)

2.5.1 Method of Moments (MM)

Method of Moments relates the sample values of the *moments to the parameters of the distribution The rth sample* moment about any arbitrary x(o) is given

$$m_r = 1/N \sum_{i=1}^{N} [x(i)-x(o)]^r \qquad (2.23)$$

where N is the size of the sample , which is an approximation to the rth population moment,

$$\int_{-\alpha}^{\alpha} (x - x_o)^r\, f(x)\, dx \qquad (2.24)$$

For any distribution, The first moment about the origin gives the

sample mean and the second moment about the mean gives the sample variance, viz ,

$$\mu \simeq \bar{x} = 1/N \sum_{i=1,N} x(i) \quad , \qquad (2.25)$$

$$\sigma^2 \simeq \text{sample var}(x) \quad \text{and} \qquad (2.26)$$

$$sx^2 = 1/(N-1) \sum_{i=1,N} [x(i)-\bar{x}]^2 \qquad (2.27)$$

$\bar{x}$ and var(x) are the sample estimates of μ and σ^2 for the theoretical distribution

Estimation by the method of moments is asymptotically efficient and the efficiency is usually smaller than unity For small samples, the estimates are significantly affected by extreme or offcontrol points that may be present in the sample

2.5.2 Method of Least Squares (MLS)

This method consists in the estimation of the parameters of the assumed distribution by minimising the sum of squares of deviations of the observed points from the fitted function Chow (1964) suggests the following procedure

If $\bar{x}$ and sx are the sample mean and standard deviation, the frequency factor K is defined as

$k = (x-\bar{x})/sx$ For any given distribution, K can ,be related to the cumulative distribution F(x) of the distribution. For a given x_i, $F(x_i)$ is estimated from plotting position formulae. For the assumed distribution, K_i is obtained from $F(x_i)$. The regression line of x on K is determined by MLS This gives estimates of sample mean $\bar{x}$ and standard deviation sx. Though this procedure is not theoretically exact, it generally gives a better overall fit than the MM and further, the estimate is not affected very much by extremely rare occurrence as in the case of MM.

A plot between the frequency factor and the variable is plotted with the variable x being on the ordinate and the frequency factor K being on the abscissa By definition, this will be a straight line. The frequency factors for different distributions are calculated as follows

a) Normal distribution ' The frequency factor can be expressed as

$K_T = (x_T - \bar{x})/sx$, where $\bar{x}$ and sx are the mean and standard deviation of the data and $K_T = Z$, which is known as the standard normal variate

For Log Normal distribution, the same procedure can be applied except that it is applied to the logarithms of the variables, and their mean and standard deviation are used

b) Extreme value distribution For Extreme value type 1 distribution, Chow (1953) derived the expression

$$K_T = -\sqrt{6}/\pi\ [\ 0.5772 + \ln\{\ln(T/T-1)\}] \qquad (2\ 28)$$

where T is the return period The value of T is calculated as $T = 1/P(x)$ where P(x) is the probability of exceedence obtained from the plotting position formula

c) Log Pearson Type 3 distribution For this distribution, the first step is to take the logarithms of the hydrologic data, $y = \log x$ Usually, logarithms to the base 10 are used The mean $\bar{y}$, standard deviation s_y and the coefficient of skewness C_s are calculated for the logarithms of the data The frequency depends on the return period T and the coefficient of skewness C_s. When $C_s = 0$, the frequency factor is equal to the standard normal variate Z When $C_s \neq 0$, K_T is approximated by Kite (1977) as

$$K_T = Z + (Z^2-1)K + 1/3\ (Z^3-6Z)K^2 - (Z^2-1)K^3 + ZK^4 + 1/3K^5 \qquad (2\ 29)$$

where $K = C_s/6$. The value of Z is the standard normal variate for given T

2.5.3 Method of Maximum Likelyhood (MLE)

Maximum likelyhood estimate is that estimate of the parameters of a distribution for which the probability of occurrence of the actual observations is a maximum

Let x(1), x(2), x(N) be the N observations and f(x), the assumed probability density function in terms of the parameters P(1), P(2), P(J) Assuming independence of events, the probability of the outcome p is given by

$$P = \sum_{i=1,N} f[x(i)] \qquad (2\ 30)$$

For this to be maximum, $\partial P/\partial P_j = 0$ for all $j = 1,2, \ldots J$. The J equations in terms of J parameters are solved to give the maximum likelyhood estimates. For the Normal distribution, these estimates are the same as the estimates by the MM For the two parameter Log Normal distribution, The MLE of the mean and the standard deviation are given by the mean and standard deviation of the

logarithm of the sample

2.6 Identification of distribution

For preliminary selection of pdf's, statistical parameters are used generally, arithmetic mean is believed to be the best central value parameter The coefficient of skewness often acts as regionalisation parameter and becomes very important when mean and standard deviation have small variation Often a combination of skew and kurtosis determines the pdf type A guide to the selection of pdf in terms of kurtosis and skewness of the sample data is given in Fig 2 3 (Roudkivi, 1979) Some criteria for the initial selection of pdf based on estimated sample statistics are presented in Table 2.1.

The following notations are used .

μ = Population mean

σ = Population standard deviation

C_s = Coefficient of skewness of population

CV = σ / μ = Coefficient of variation

C_k = Coefficient of kurtosis

Selection of pdf largely depends upon the type of hydrologic variables, viz , flood, drought, rainfall, etc for low floods or droughts, Extreme value type 3 or Pearson type 3 distributions (Kite, 1977) will be suitable For floods, Log Normal, Pearson type 3, Extreme value type 1 and Extreme value type 3 distributions are suitable For rainfall, Extreme value type 1 distribution is preferable (Chow, 1964) Similarly for Exceedence series, Negative Exponential distribution is well suited

2.7 Outliers

An outlier in a set of may be defined as an observation (or a subset of observations) which appears to be inconsistent with the remainder of that set of data In picking out an observation (or a set of observations) as outliers, the main problem is in deciding whether or not some observations are genuine members of the main population If they are not, then the attempts made to draw inferences about that population may be

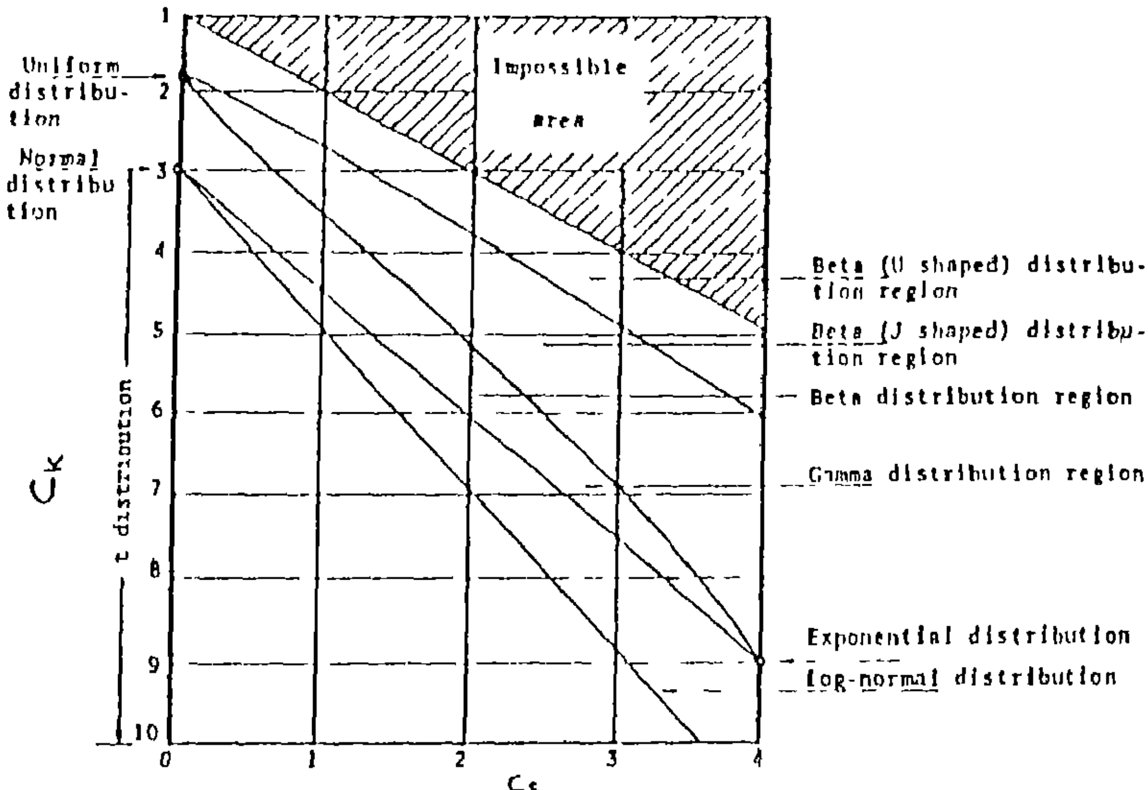


Fig 2 3 Regions in (C_s C_k) plane for various distributions, after E S Pearson From Hahn and Shapiro (1967).

Table 2.1

Theortical and Heuristic Parameters Ranges
for Different Frequencies Distributions

case nos.	Theortically	Heuristically	Suggestion
1.	$C_s = 0.0$	$-0.5 < C_s < 0.5$	
	$CV < 0.4$	same	
	$C_k > 1.0$	same	
	and		
	i. if $C_k < 2.0$	if $C_k < 2.25$	Uniform distribution
	ii. if $C_k = 3.0$	if $2.0 < C_k < 4.0$	Normal distribution
	iii. if $C_k > 3.0$	if $4.0 < C_k < 10.0$	t distribution
2.	$C_s < 0.0$	$C_s \leq -0.5$	Pearson type III or Lognormal type III can be fitted. Priority may be given to Pearson type III on the basis of parsimony of parameters.
3.	$C_s > 0.0$	$C_s > 0.5$	
	i. $C_k < (1.25C_s + 1.0)$ or $C_k < 1.0$	same	No distribution (impossible region)
	ii. $C_k = 2C_s + 3.0$	$1.25C_s + 1.3 < C_k < 2.25C_s +$	Lognormal II distribution

contd

iii. $C_s = 1.14$ and $C_k = 5.4$	$0.64 < C_s < 1.64$ $4.5 < C_k < 6.5$	Extreme value type I distribution
iv. $C_k = 3.0+6CV$ Where, $CV = 1/\sqrt{N}$	$2+6CV < C_k < 4+6CV$ Where, $\frac{1}{\sqrt{N}} -0.2<CV<\frac{1}{\sqrt{N}} +0.2$	Gamma distribution
v. $C_s = 4.0$ $C_k = 9.0$	$3.5 < C_s < 4.5$ $8 < C_k < 10$	Exponential distribution
vi. $(1.0625C_s+1.75)< C_k<(1.8125C_s+1.75)$	same	Beta (j) distribution
vii. $(C_s+1.0) < C_k<(1.0625C_s+1.75)$	same	Beta (u) distribution
viii. $(1.250C_s+3.0) < C_k<(1.8125C_s+1.75)$	same	Beta or Gamma distribution

Note : Alternatively

1. Input data can be log transformed to a Y series and then transformed data can be tested for Normal,Pearson type III or External type I disributions.

misleading Hence, it becomes necessary to know the basic distinction between *Extreme observations*, *Outliers* and *Contaminants*

2.7.1 Difference between Extremes, Outliers and Contaminants

Suppose there is a random univariate sample of size n , x_1, x_2, ...,x_n from a distribution whose form is denoted by F. The ordered sample (from smallest to largest) is $x_{(1)}$, $x_{(2)}$, . . ,$x_{(n)}$ The observation $x_{(1)}$ and $x_{(n)}$ are the sample *Extremes* To declare either of them as an *Outlier* depends on how they appear in relation to the postulated model F In Fig 2 4(a), neither $x_{(1)}$ nor $x_{(n)}$ appears to be outlying. In Fig 2 4(b), $x_{(n)}$ is an upper outlier and $x_{(1)}$ also gives some cause for concern. $x_{(1)}$ may be declared as an lower outlier (so that $x_{(1)}$, $x_{(n)}$ is an outlier pair $x_{(n-1)}$, $x_{(n)}$ may be declared as an upper outlier pair) Hence, extreme values may or may not be outliers Any outliers, however, are always extreme (or relatively extreme) values in the sample

Suppose that not all the observations come from the distribution F, but one or two come from a distribution G which has slipped upward relative to F (i e it has a larger mean) The observations from G are termed *Contaminants* Such contaminants may appear as extremes but need not do so Fig 2 4(c) shows two contaminants (indicated by), one of which is upper extreme, the other is in the midst of the sample But, $x_{(n)}$ whilst an extreme and a contaminant, is not an outlier In Fig 2 4(d), however, a non-extreme contaminant is seen which is nonetheless outlying as one of the outlier pair $x_{(n-1)}$,$x_{(n)}$ If the number of observations from distribution G is not negligible in terms of the total sample size n, it may be necessary to treat the distribution as a mixed distribution, say two distribution

Here outliers may or may not be contaminants, and contaminants may or may not be outliers There may not be any method to know whether or not any observation is a contaminant So attention is concentrated on outliers as the possible manifestation of contamination

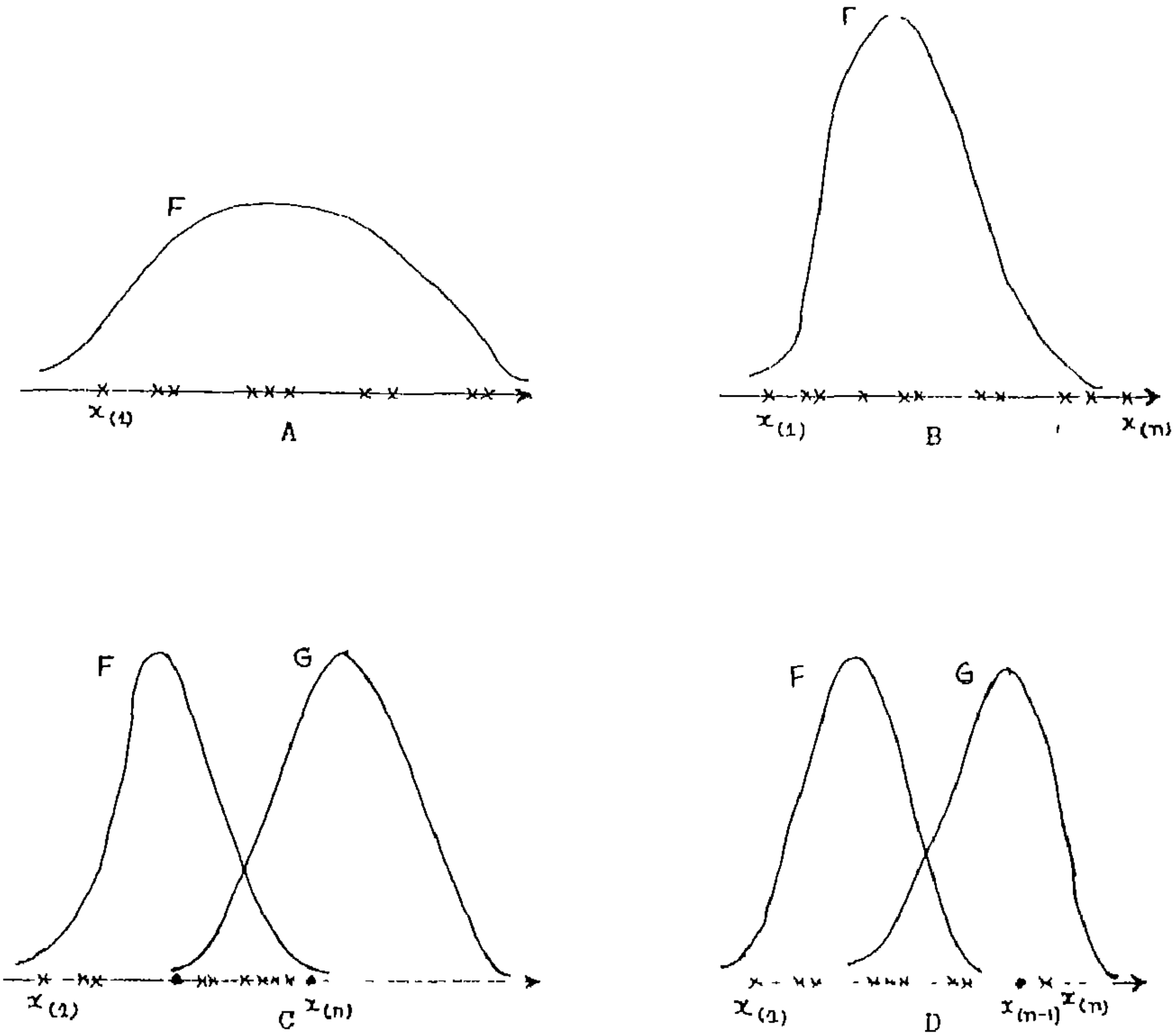


Fig 2 4 Extremes, Outliers and Contaminants

2.7.2 Origin of Outlier

In taking observations, different sources of variability can be encountered Three of those that can be distinguished are

a) Inherent variability This is the expression of the way in which observations vary over the population; such variation is a natural feature of the population It is uncontrollable and reflects the distributional properties of a correct basic model describing the generation of the data Thus, for example, measurements of heights of men will reflect the amount of variability indigenous to the population (and this may well be reasonably modelled by a Normal distribution)

b) Measurement error Often, physical measurements are taken on members of a population under study Inadequacies in the measuring instrument superimpose a further degree of variability on the inherent factor. The rounding off of obtained values, or mistakes in recording, compound the measurement error they are part of it Some control of this type of variability is possible.

c) Execution error A further source of variability arises in the imperfect collection of data A biased sample may be chosen or individuals included are not truly representative of the population that is being aimed to sample Sensible precautions may reduce such variability but one may not be aware of the execution errors and sometimes it may be appropriate to change the basic population model to encompass the prospect of anomalous sample members

2.7.3 Tests for Outliers

There are a number of approaches to identify and test outliers (Barnett V and Lewis.T, 1983) Some simple approaches are given below

i) Normal and Log Normal distributions The following equation can be used to detect high outliers

$$Y_H = \bar{y} + K_n s_y \qquad (2.31)$$

where Y_H is the high outlier threshold and K_n is in terms of sample size as given in Table 2.2 K_n values in Table 2.2 are used in the one sided tests that detect outliers at the 10 % level of significance in normally distributed data If the values in the

Table 2.2
Outlier test Kn values

Sample size n	Kn	Sample size n	Kn	Sample size n	Kn	Sample size n	Kn
10	2.036	24	2.467	38	2.661	60	2.837
11	2.088	25	2.486	39	2.671	65	2.866
12	2.134	26	2.502	40	2.682	70	2.893
13	2.175	27	2.519	41	2.692	75	2.917
14	2.213	28	2.534	42	2.700	80	2.940
15	2.247	29	2.549	43	2.710	85	2.961
16	2.279	30	2.563	44	2.719	90	2.981
17	2.309	31	2.577	45	2.727	95	3.000
18	2.335	32	2.591	46	2.736	100	3.017
19	2.361	33	2.604	47	2.744	110	3.049
20	2.385	34	2.616	48	2.753	120	3.078
21	2.408	35	2.628	49	2.760	130	3.104
22	2.429	36	2.639	50	2.768	140	3.129
23	2.448	37	2.650	55	2.804		

Source. U.S. Water resources Council, 1981. This table contains one sided 10-percent significance level Kn values for the normal distribution

sample are greater than Y_H in Eq 2 31 , then they are considered high outliers

A similar equation can be used to detect low outliers ,

$$Y_L = \bar{y} - K_n\ s_y \qquad (2\ 32)$$

where Y_L is the low outlier threshold in log units, $\bar{y}$ and s_y are the mean and standard deviation of the data

According to the Water Resources Council (1981), if the station skew is greater than + Ø 4, tests for high outliers are considered first, if the station skew is less than - Ø 4, tests for low outliers are considered first Where the station skew is between + Ø 4 and -Ø 4, tests for both high and low outliers should be applied before eliminating any outliers from the data set

For data distributed log normally, the log transformation transforms the data to Normal and the above procedure can then be applied to the logarithms

ii) Pearson and Log Pearson Type 3 distributions If the original distribution is Pearson distribution, it can be transformed to a Normal distribution by the Hilferty - Wilson transformation and then tested for outliers in the Normal distribution The procedure consists in first standardising the raw series so that the resulting series has nearly zero mean and unit standard deviation

Let x_i be the Pearson Type III standard deviate obtained from the standardised data with skew coefficient g Then, the corresponding normal(Ø,1) deviate t_i can be obtained by the transformation

$$t_i = 6/g\ [(g\ x_i/2 + 1)^{1/3} - 1) + g/6 \qquad (2\ 33)$$

In case the probability distribution is a log Pearson Type III distribution, then the log transformation is used followed by standardisation and then Hilferty - Wilson transformation which transforms the distribution to a nearly Normal distribution

iii) Extreme value Type 1 and Type 3 distributions Extreme value Type I distribution is first transformed to the Exponential distribution and then tested for outliers in the Exponential distribution.

If X has a Gumbel greatest value distribution, the transformed random variable Y = exp (-x/b) has an Exponential distribution with origin zero and scale parameter exp (-a/b) where

$$b = s_x \sqrt{6} / \pi \quad \text{and} \tag{2.34}$$

$$a - \bar{x} - Ø\ 5772\ b \tag{2 35}$$

So, if the value of b is known, the discordancy of an outlier or a set of outliers in a sample from the X-distribution can be tested by transforming each observed value x_i to y_i = exp $(-x_i/b)$ and using on the y 's a discordancy test for an exponential sample with origin zero

In this transformation, an upper outlier $x_{(n)}$ in the X-sample converts to a lower outlier $y_{(1)}$ in the Y-sample and vice-versa

To test for discordancy in the exponential sample, a recursive procedure is adopted A test is used to identify a single upper outlier or lower outlier and is used recursively until all the outliers have been identified

For high outliers, the test statistic $(x_n / \Sigma x_j)$ is compared with the 5 % critical values given in Table 2 3 and for low outliers, the test statistic $(x_1 / \Sigma x_j)$ is compared with the 5 % critical values given in Table 2 4, for a particular length of data n The observed value is declared discordant if the test statistic is greater than the critical value in the case of high outliers and less than the critical value in the case of low outliers

The sample data are now retransformed to the Exponential distribution using the transformation

$$x_i = -b \log s_y \quad \text{where} \tag{2.36}$$

$$b = s_x \sqrt{6} / \pi \tag{2.37}$$

For the distribution of Extreme value Type III the log transformation is used first, followed by the above procedure which transforms it to exponential distribution and then tested for outliers

2.8 Test for goodness of fit

The theoretical distribution that is fit to the given sample data can be validated by the goodness of fit tests. These

Table 2.3 Critical values for 5% tests of discordancy for an upper outlier in a gamma sample using the ratio $x_{(n)} / \sum x_j$ as test statistic.

(Barnett and Lewis ; 1983)

5% critical values

n \ r	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	8	18	72	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.4 Critical values for 5% and 1% tests of discordancy for a lower outlier in an exponential sample, using $x_{(1)}/\sum x_i$ as test statistic. Values of the statistic *lower* than the critical value are significant

n	5%	1%
3	0.00844	0.00167
4	0.00424	$0.0^{3}836$
5	0.00255	$0.0^{3}502$
6	0.00170	$0.0^{3}335$
7	0.00122	$0.0^{3}239$
8	$0.0^{3}913$	$0.0^{3}179$
9	$0.0^{3}710$	$0.0^{3}140$
10	$0.0^{3}568$	$0.0^{3}112$
12	$0.0^{3}388$	$0.0^{4}761$
14	$0.0^{3}281$	$0.0^{4}552$
16	$0.0^{3}213$	$0.0^{4}419$
18	$0.0^{3}167$	$0.0^{4}328$
20	$0.0^{3}135$	$0.0^{4}264$
30	$0.0^{4}589$	$0.0^{4}116$
40	$0.0^{4}329$	$0.0^{5}644$
50	$0.0^{4}209$	$0.0^{5}410$
100	$0.0^{5}518$	$0.0^{5}102$

n – number of observations

tests indicate the goodness of fit for a certain distribution at a particular confidence level The test used most frequently is the Chi Square test as described below

The sample space is divided into I mutually exclusive classes with a class frequency of 5 or more Let $p(i)$ be the probability that the variable belongs to the i th class for the assumed distribution If $x(i)$ and $x(i+1)$ are the limits of the i-th class interval, then

$$p(i) = F[x(i+1)] - F[x(i)] \qquad (2\ 38)$$

Let $f(i)$ be the observed frequency of the sample from the i-th group If N is the total number of samples, the chi square statistic is given by

$$\chi^2 = \sum_{i=1}^{I} \frac{[f(i) - N(p(i))]^2}{N\ p(i)} \qquad (2.39)$$

If J is the number of parameters estimated, then theoretically χ^2 has a distribution with (I-J-1) degrees of freedom Let $\chi^2(\alpha)$ denote the value of χ^2 at α % confidence level for the above degree of freedom as obtained from the tables If the calculated χ^2 is greater than the theoretical value, then the sample deviates significantly from the assumed distribution at the given confidence level and the fit is rejected If it is less, then the fit is accepted.

2.9 Confidence Bands

Statistical estimates are often presented with a range or confidence interval, within which the true value can reasonably be expected to lie The size of the confidence interval depends on the *Confidence level* β The upper and lower values of the confidence interval are called *Confidence limits*

Corresponding to the confidence level β is a significance level α, given by $\alpha = (1-\beta)/2$. For estimating the event magnitude for the return period T, the upper limit $U_{T,\alpha}$ and lower limit $L_{T,\alpha}$ for a Normal distribution are given by

$$U_{T,\alpha} = \bar{y} + s_y\ K^{U}_{T,\alpha} \qquad \text{and} \qquad (2.40)$$

$$L_{T,\alpha} = \bar{y} + s_y\ K^{L}_{T,\alpha} \qquad \text{where} \qquad (2.41)$$

$K^U_{T,\alpha}$ and $K^L_{T,\alpha}$ are the upper and lower confidence limit factors, which can be used for the Normal distribution and Pearson Type 3 distribution Approximate values for these factors are given by

$$K^U_{T,\alpha} = (K_T + \sqrt{K_T^2 - ab})/a \quad \text{and} \qquad (2.42)$$

$$K^L_{T,\alpha} = (K_T - \sqrt{K_T^2 - ab})/a \quad \text{in which} \qquad (2.43)$$

$$a = 1 - Z\alpha^2/2(n-1) \quad \text{and} \qquad (2.44)$$

$$b = K_T^2 - Z\alpha^2/n \qquad (2.45)$$

The quantity $Z\alpha$ is the standard normal variable with exceedence probability α The same factors are used to construct approximate confidence limits for Extreme value Type 1 dsitribution

2.10 Expert Systems in Frequency Analysis

There is no general agreement among hydrologists on the specific choice of any particular theoretical distribution for frequency analysis of a given hydrologic variable at a given site. Frequency analysis, being a data dependent technique is subject to many constraints, limitations and assumptions

The data are assumed to be consistent, homogeneous and independent which in actual practice need not be true The maximum floods are seldom, if ever, measured correctly due to uncertainity in occurrence time and lack of proper gauging facilities as well as rating curves at the proper time and place. As a result they are estimated from the extrapolation of rating curves and other techniques. These estimates causo inconsistency in data with possible large errors There are many more constraints like tho variable type, sample size, outliers in data, nonstationarity with respect to the process involved etc which have to be taken care of.

These problems, if not dealt with proper attention and with an intelligent knowledge rich heuristic approach, will often lead to misleading results Hence, domain specific human expertise can be effectively used to tackle these problems. The ES may take care of these problems to a large extent and fit a suitable theoretical probability distribution function to observed data

CHAPTER 3

EXPERT SYSTEM

3.1 Introduction

Human experts in any field are frequently in great demand and are also generally in short supply AI presents a solution to such a problem through an expert system (ES) which is a computing system capable of representing and reasoning about some knowledge rich domain with a view to solving problems and giving advice (Jackson, 1986) It is also known as Knowledge based expert system (KBES)

Gasching (1981) defines KBES as ' interactive computer program incorporating judgement, experience, rules of thumb, intuition and other expertise to provide knowledgeable advice about a variety of tasks ' So, ES 's act as an intelligent assistant to human a expert and also provide assistance to people who otherwise might not have access to expert advice

3.2 Building Expert System

3.2.1 Architecture of KBES

The KBES generally has four principal components (Fig - 3 1) They are a knowledge base, working memory, inference engine and a user interface As KBES vary in design, they may have other components also for eg , graphics, system analysis and other software

a) Knowledge base A Knowledge base contains both declarative knowledge (facts about objects, events and situations) and procedural knowledge (information about courses of action) which may be scientific, analytic or heuristic rules (Fig 3 2) Although many knowledge representation techniques have been used in ES, the most prevalent form of knowledge representation currently used in ES is the ' rule based production system approach The rules have generally two parts, conditions and actions The rules are fired when the conditions are matched with the facts. The actions can be for processing instructions or to control instructions. The rules may include meatrules which are rules about rules

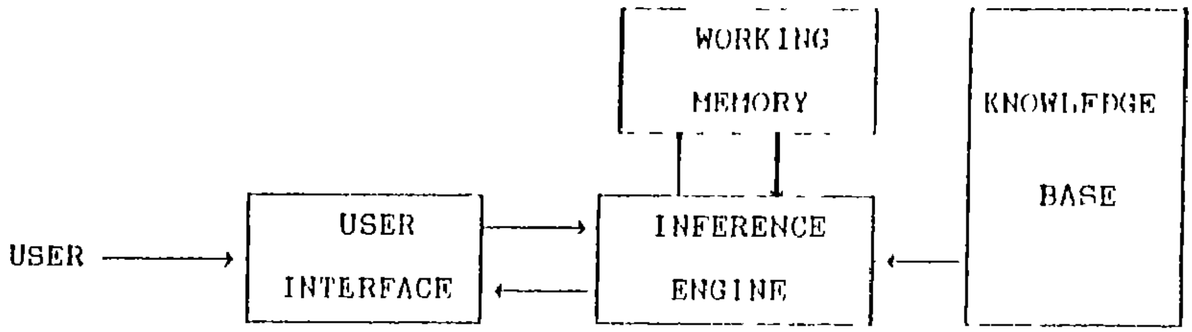


Fig 3 1 ARCHITECTURE OF A TYPICAL ES

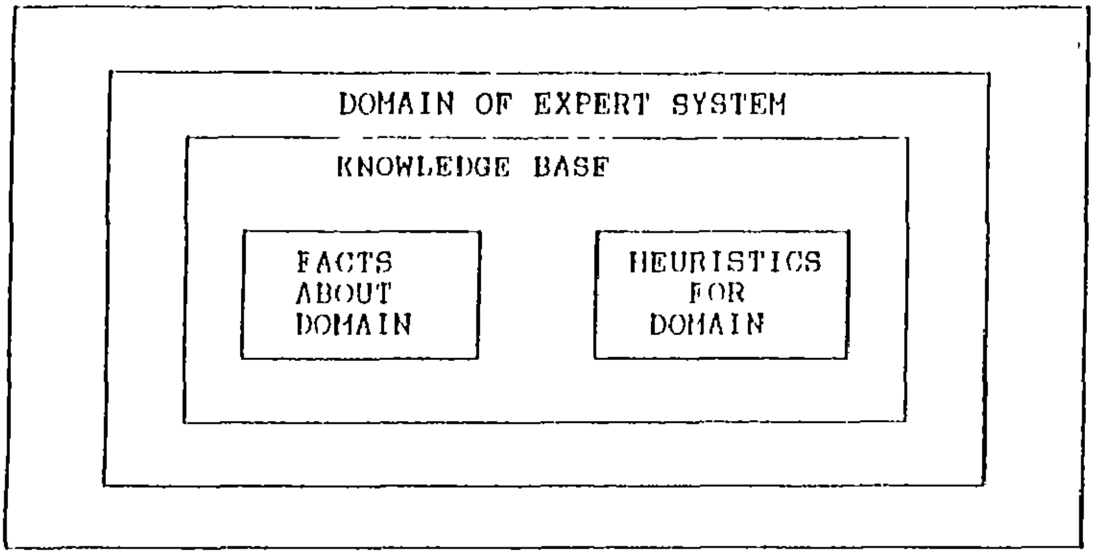


Fig 3 2 : THE COMPONENTS OF THE KNOWLEDGE BASE OF AN ES (Schank and Childers, 1984)

b) Working memory It is the current active set of the knowledge base and may include a knowledge management module

c) Inference engine It is the most crucial component of ES since it matches and manipulates the database for problem solving. It is the inference mechanism which also provides justification for the ES Three formal approaches used in this case are production rules, structured objects and predicate logic. Production rules consist of a rule set, a rule interpreter which specifies when and how to apply the rules and working memory that holds data, goals or intermediate results Structured objects use vector representation of essential and accidental properties Predicate logic uses prepositional and predicate calculus

d) User interface It is the communication module which provides bidirectional exchange of information between the user and system

3.2.2 ES technique

The order of execution of the rules and/or procedures in an ES is governed by the inference engine in terms of the problem solving strategy used Maher (1986) considers two approaches

a) The derivation approach It involves deriving a solution that is most appropriate to the problem from a list of predefined solutions stored in the knowledge base of the ES It includes *forward chaining* (or goal driven control strategy) , *backward chaining* (or data driven strategy) and a hybrid strategy combining both these strategies Forward chaining works from an initial state of known facts to the goal state and backward chaining works from a hypothetical goal state to the facts perhaps in terms of subgoals The subgoals are preconditions for the goal stated If the hypothesis is not supported by facts it tests for another goal state and so on in a predefined order of goals

b) The formation approach It involves forming a solution from eligible solution components stored in the knowledge base It includes problem reduction (into subprograms), plan-generate-test (which generates all possible solutions, prunes inconsistent solutions and tests the remaining solutions), and agenda control In agenda control, a priority rating to each task in the agenda is

assigned and the tasks are performed according to the assigned priority

These techniques may be combined with other techniques for hierarchical planning, least commitment backtracking and constraint handling (Maher, 1986) Some other techniques available include inductive inference, metareasoning, ill structured problem and data handling etc

3.2.3 Developing an ES

Developing as ES ia a time consuming team work Particularly, for developing a sophisticated ES , an intensive and coherent effort is required Knowledge engineers and domain experts work together to design an ES The knowledge engineer develops the expert system and the domain expert provides the information for the knowledge base Hayes-Roth et al (1983) have identified five sequential stages in the development of ES, shown in Fig 3.3 Each stage is iterative in nature Some of the stages are shown Fig 3 4 They are self explanatory

3.2.4 ES Tools

A wide variety of development tools and environments are available for ES These tools can be one of the general purpose programming languages or an ES shell An ES shell is a set of ES development programs containing no knowledge about a problem, but can be taught in a particular field or other They contain all the modules required for ES Filling of their hollow knowledge bases makes them knowledge based expert system Broadly, the ES tools are catagorised As

1 Programming languages like PASCAL, C ,
2. AI based extrapolatory programing languages like LISP and PROLOG ,
3. ES shells like VIDHI which may be based on item 2 ,
4 High level ES programming environments like OPS5, ART, KES, NEXPERT, PC PLUS, RULEMASTER, CLIPS etc , and
5 Mixed programming environments which allow the programmer to mix programming languages as in item 1 and item 2 with high level ES programming environment as in item 4

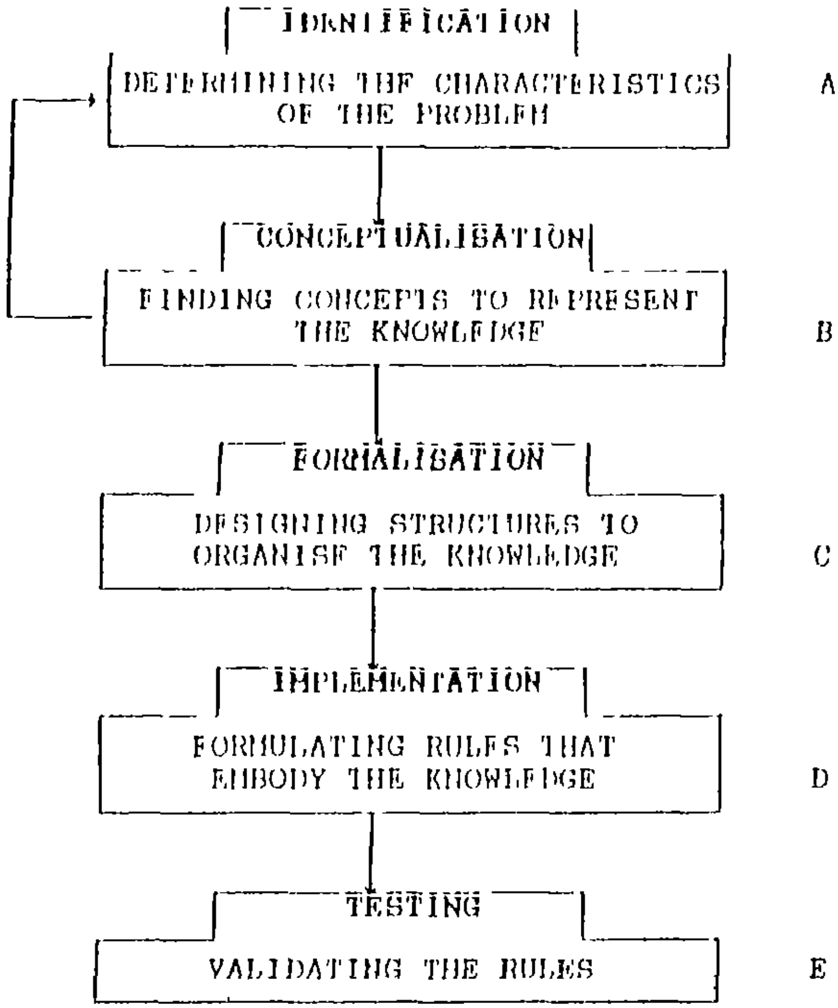


Fig 3 3: FIVE STAGES OF ES DEVELOPMENT
(Adapted from Hayes - Roth et al, 1983)

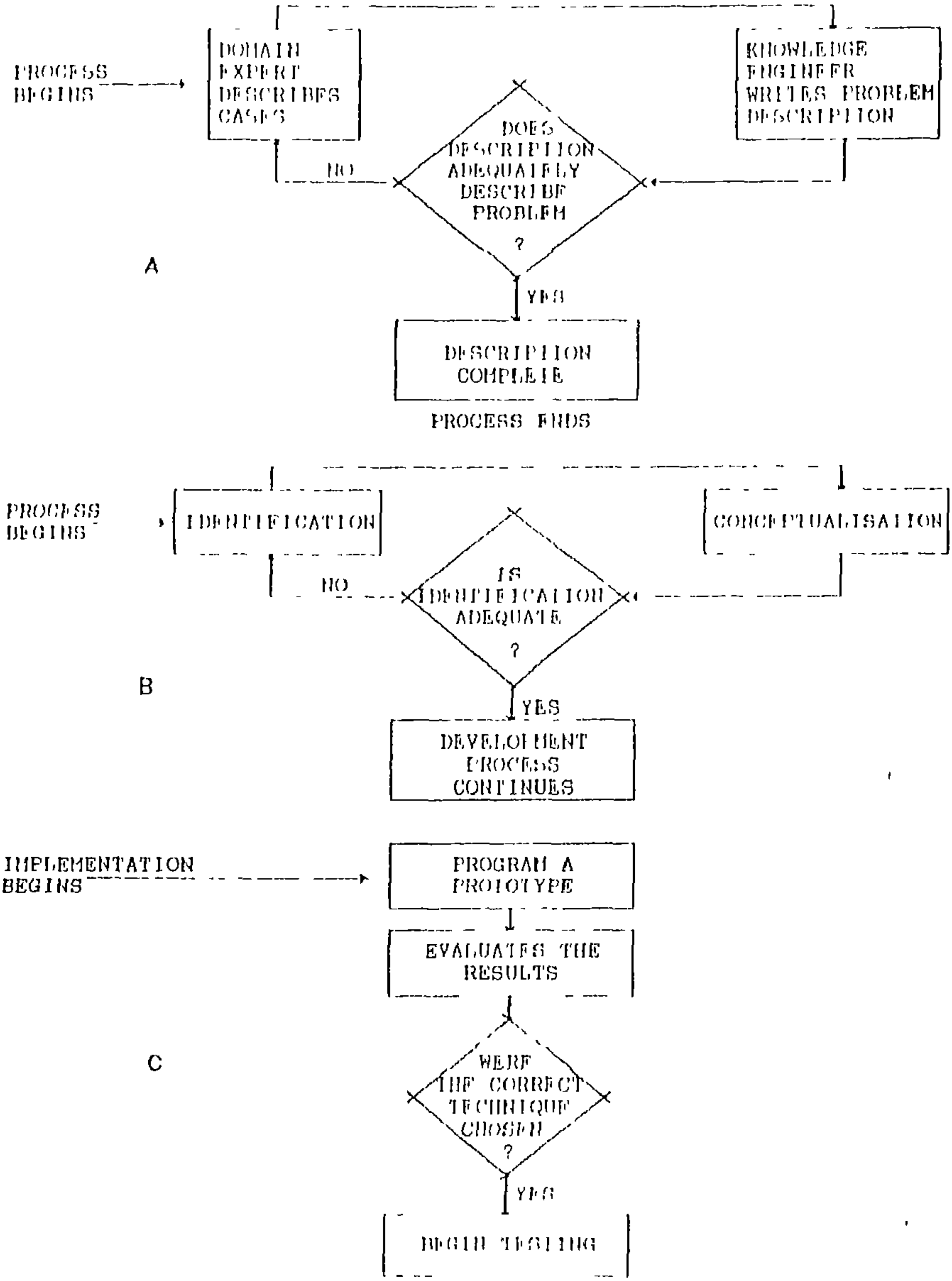


Fig. 3.4. ITERATIVE STAGES OF ES DEVELOPMENT

3.3 CLIPS as an ES Shell

3.3.1 Introduction

CLIPS ('C' Language Integrated Production System) is a rule based forward chaining mixed programming expert system shell developed by National Aeronautics and Space Administration (NASA), Houston, Texas The source program of CLIPS (version 4 3) is written in C language THe primary representation methodology is a forward chaining rule language based on Rete algorithm (Forge, 1982) for multiple pattern matching and conflict resolution CLIPS is a versatile software in that it can be incorporated with any other language like FORTRAN and C CLIPS can be used for developing a KBES either as an Interactive executable element or as an Embedded executable element

3.3.2 Knowledge representation

a) Facts : A Fact represents a piece of information and is placed in a current list of facts, the *fact list* It acts as a pattern for matching the conditions of a rule in order to fire that particular rule. Facts can be added (asserted) in two ways, viz., prior to execution and by the action of a rule firing A fact can be a number, a word, or a string Facts are defined using *deffacts* construct.

b) Rule The primary method of representing knowledge in CLIPS is a *Rule* A rule is a collection of actions to be taken if the conditions are met The conditions are patterns which act as constraints and also provide a way to describe how to solve a problem. Rules are defined using the *defrule* construct

A rule is made up of a LHS part and RHS part The LHS of a rule is made up of a series of one or more patterns which represent the condition element of the rule In RHS, a list of the actions to be performed when LHS of the rule is satisfied is given The arrow (=>) separates RHS from LHS Two important features of the RHS actions are the use of ' if then-else ' structure and 'while ' structure

Further details on facts and rules available in CLIPS are given in CLIPS manual (Culbert, 1987)

3.3.3 Inference Engine

The inference engine of CLIPS is rule based and as indicated earlier (Sec 3.2.1) works on a forward chaining inference mechanism It is based on Rete algorithm (Forge,1984) for multiple pattern matching and conflict resolution

3.3.4 Cycle of Execution

In CLIPS, the starting and stopping points are not explicitly defined The inference engine applies the knowledge (rules) to the data (fact). The cycle of execution is as follows :

1 The knowledge base is examined to see if the conditions of any rule has been met

2. All rules whose conditions are currently met are activated and placed on the agenda (a stack) Rules having higher priority are kept on the top of the stack and are activated before a new rule Rules having lower or equal priority remain below the new rule

3. The top rule on the agenda is selected and its RHS actions executed.

As a result of RHS actions new rules can be activated and/or deactivated

This cycle is repeated until all rules that can fire have done so, or until a so called rule limit is reached.

3.4 Embedded application of CLIPS

CLIPS has an added advantage of being integrated with external functions and/or C, FORTRAN or Ada language programs. This capability makes its application very flexible and more suited to engineering applications where numerical computations are of great importance, and where computer programs in any of the languages are already available

3.4.1 External Function

An external function defined by the user for his specific use in problem solving can be either in C or in the language within which CLIPS is being embedded An external function can be used in both the LHS and RHS of the rules Data can be passed to and from them

The other method of passing data, asserting a new fact directly into the CLIPS fact list, is done by calling the C function assert

CLIPS provides some more advanced interface functions in which passing known variable types, accessing multifield variables, and building facts by scratch (to assert a large number of facts) are possible (Culbert, 1987)

3.4.2 Embedded application

a) General CLIPS is designed to be embedded within other programs The embedding program can be a C language program or FORTRAN program or a Ada program In each case a main program is provided by the user which calls CLIPS like any other subroutine. The basic changes which are to made to access CLIPS from main program in a different language is available (Culbert,1987 ; Mishra,1990)

b) FORTRAN - CLIPS interaction

For complete language mixing four basic capabilities are needed

1) A program in another language may be used as the main program and CLIPS can be called as needed for reasoning,

2) Facts can be asserted into CLIPS from other languages,

3) CLIPS may call other functions written in any language from the RHS of a rule and may pass parameters to the function , and

4) In languages which can provide a meaningful return value, external functions may be called from the LHS of a rule (i e , used as predicate function).

The main program written in FORTRAN initialises CLIPS, loads the rule files, resets the process,asserts the facts and runs the program by calling specialised CLIPS functions The functions are , init_clips, load_rules, reset_clips, assert and run_clips in a sequence In order to load a rule file or to assert a fact the FORTRAN-strings are to be converted to C-strings by calling a function storec Similarly for reverse action, i e , to pass a parameter from CLIPS to FORTRAN, the C-string is to be converted into FORTRAN by calling a function *loadc*

Facts can be asserted to CLIPS from FORTRAN either as constraints or as variables

These functions are defined under 'usrfuncs' either in file 'Main.c' (in interactive mode) or any other file (in embedded application) Within usrfuncs a call should be made to the define-function routine for every function about which the user wants CLIPS to know User defined functions are searched before system functions and if it matches with one of the defined functions already provided, the user function will be executed in its place

a) Passing variable from CLIPS to external function

CLIPS actually calls the function without any arguments, though they are listed directly following a function name inside CLIPS rules Instead the parameters are stored internally by CLIPS and can be accessed by calling the functions .

```
int num_arg() ,
char rstring(arg) ,
float rfloat(arg) ,
int runknown(arg) ,
int arg ,
```

A call to num_arg will return an integer telling how many arguments the function was called with

A call to rstring returns a character pointer and rfloat returns a floating point number The parameters have to be requested one at a time by specifying the parameter position number as the argument to rstring or rfloat If the type of the argument is unknown, runknown can be called to determine the type.

b) Passing data from external function to CLIPS

An external function can pass data into CLIPS in two ways It can return a value or can assert a new fact directly into the CLIPS fact_list If the external function is to be used as predicate, it must return a floating point number otherwise it can be a character, integer, word or unknown Return values can be used as predicates, bound to variables or captured via pattern recognition The return values do not have to be captured, but must be defined in CLIPS, and all external functions must return a value

3.5 Structure of FACHVES program

3.5.1 Introduction

Frequency analysis of continuous hydrologic variables embedding expert system (FACHVES) is an interactive fortran program with an embedded expert system The program consists of a main program, subroutines, ES interface and a rule based knowledge base

3.5.2 Environment

FACHVES programs have been developed , implemented and tested in VAX-VMS environment .However it can be implemented on HP 9000-800 UNIX environment on IBM PC with slight modifications

3.5.3 Main Program

The main program of FACHVES is a menu based program consisting of one master menu and eight submenus

The master menu can call any of the eight submenus depending upon the option of the user The submenus are basically for preliminary analysis, method of analysis, choice of distribution, transformations, goodness of fit test, plotting of results and ES interface Each submenu consists of several options and depending upon user's choice, the desired subroutine can be called However, the user is supposed to create an input file or give input data interactively before calling the master menu for the first time Options for seasonal data analysis is displayed in the master menu only when the data is seasonal

3.5.4 Subroutines

In FACHVES general subroutines already available in FORTRAN for statistical analysis and some subroutines, eg , for two step power transformation have been incorporated They can be called either from main program or from other subroutines, as per requirement The subroutines are basically for

a) fitting different distributions

b) applying different goodness of fit tests

c) transforming the data using different transformations

d) using different methods of analysis

e) linking FORTRAN program to CLIPS and

f) converting strings from FORTRAN to C and vice versa

The subroutines are programmed in such a way that they perform one or a combination of the above functions The limitations in using a subroutine, if any, are conveyed to the user interactively during execution

One of the subroutines PTHELP acts as ES interface and it has been incorporated as an optional calling subroutine from FORTRAN main program The subroutine PTHELP links the main program with the CLIPS function calls and string conversion function STOREC The suggestions given by ES is converted to FORTRAN strings by LOADC function and stored in a subroutine ADVICE which subsequently passes them to the main program

3.5.5 Knowledge Base

The main function of the ES is to supplement the computational procedures specified in the main program with a number of advices for choosing the right alternatives for analysing the data Without the ES knowledge base (KB) similar task could have been accomplished, however, only by going through an exhaustive process of all the available alternatives The ES knowledge base helps to eliminate some of the nonfeasible alternatives of the pdf fitting exercise It also incorporates subjective judgements inbuilt in the KB, or provided by the user, to choose the right path of evaluation.

The ES is embedded in the main program and accessed by it during execution to accomplish the following tasks

i) provide an interface for user supplied information in the decision making process,

ii) transfer appropriate parameters to the FORTRAN main program, based on the advice of the ES,

iii) advice on the adequacy of the data for statistical analysis,

iv) advice on the initial choice of distribution functions based on estimated values of statistical parameters,

v) advice on the suitable choice of transformations based on the estimated values of statistical parameters, and

vi) methods of dealing with outliers in statistical data

The rule based ES's knowledge base consists of a set of rules that reflects expert knowledge in this field It also provides a vehicle for incorporating subjective judgement in the selection of procedures for statistical analysis while guiding the user through various stages of decision making The flow chart showing the various components and functions of the ES knowledge base is shown in Fig 3 5

The knowledge base is accessed during the execution of the main program The type of inputs provided to the knowledge base directly from the main program includes the estimated values of the statistical parameters of the raw hydrologic data, and the length of the record Based on these inputs, the ES guides the user through a series of steps, where existing patterns (rules) in the KB, are matched with the patterns constructed through user input and given input information Matching or non-matching of these patterns leads to either a set of parameters or control variables being passed on to the main program or the testing of a subsequent rule

Once all the existing rules compatible with the facts asserted by the user or generated by the actions of rules have been tested, all the parameters and control variables as specified by the ES are transferred to the main program Also, a list of suggestions on the subsequent procedures of statistical analysis are displayed by the ES, to help the user Whenever the user needs some advice the ES knowledge base can be accessed by the user from the main program by using either the master menu or the submenu preliminary analysis

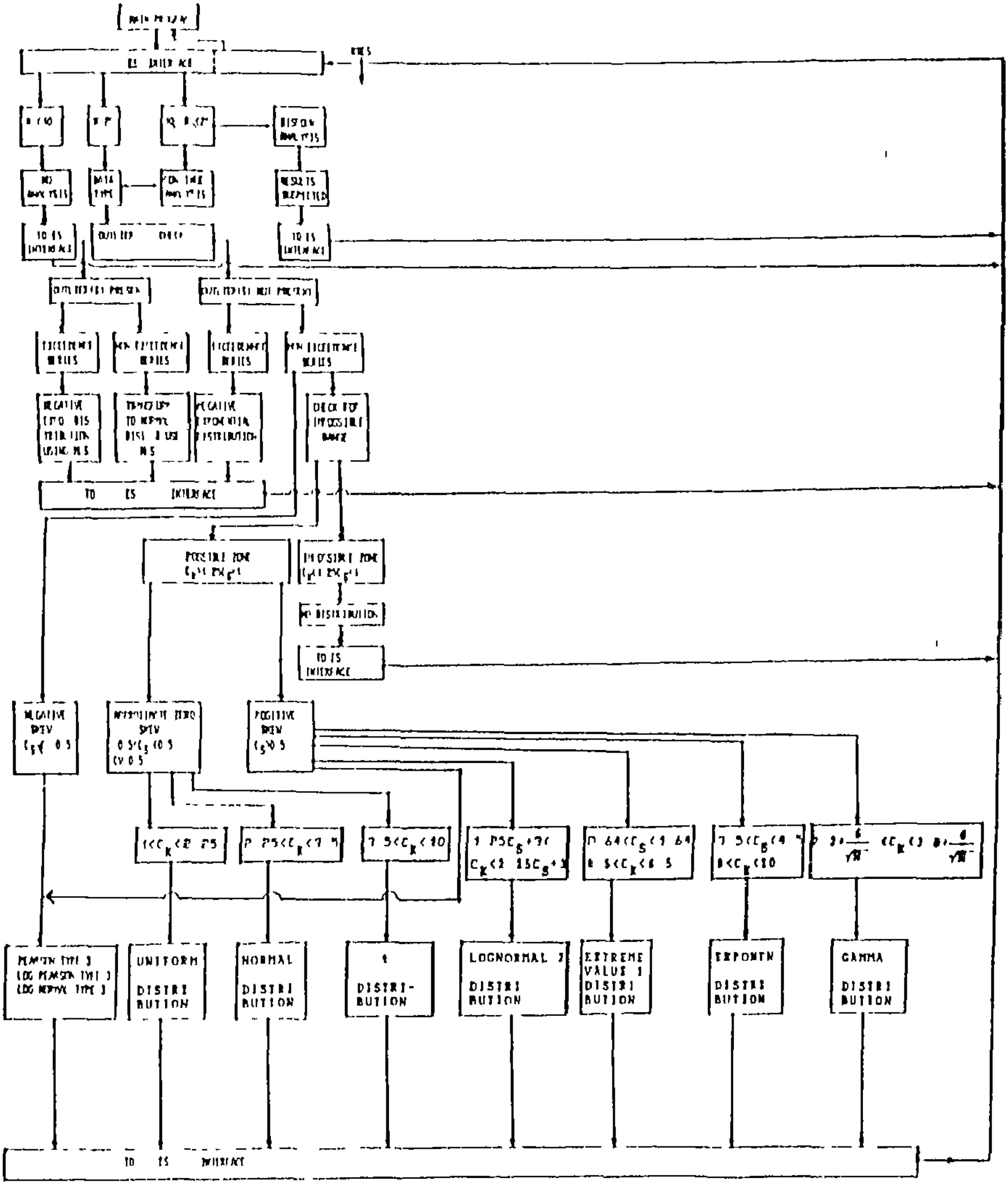


Fig 3 5 Flowchart of Knowledge Base of ES

CHAPTER 4

DESCRIPTION OF FACHVES Ver 2.0

4.1 Modifications to FACHVES

FACHVES indicated a general structure and included some programs for analysis A number of modifications and additions were made to the program and they are given below

4.1.1 Plotting of results

a) Histogram for the data

A Histogram is drawn for the given data with the frequency on the ordinate and the class interval on the abscissa The given data have been divided into eight class intervals VAX in built subroutines like Draw line, Draw rectangle etc have been used to draw the histogram The histogram comprises of a series of combinations of rectangles and straight lines For example, the routine

SMG$DRAW LINE (DISPLAY1,18,7,18,60)

draws a horizontal line on the screen from 7 th column to the 60 th column in the 18 th row

Similarly, the routine

SMG$DRAW RECTANGLE (DISPLAY1,R1,C1,R2,C2)

where R1, C1, correspond to the co ordinates of the top left hand corner of the rectangle and R2, C2 to the bottom right corner of the rectangle, draws a rectangle on the screen

b) Cumulative distribution function

The Cumulative distribution function for different distributions are drawn on the screen using a plotting subroutine The data points are plotted row wise on the screen

The cumulative distribution function is calculated from the plotting positions The plotting positions are first calculated for different distributions Most plotting position formulae are represented by the following form

$$P(X \geq X_m) = (m-b) / (n+1-2b) \qquad (4\ 1)$$

where b is a parameter, n is the total number of data values and m is the rank of an observation when the data are arranged in the descending order The value of b varies for different

distributions For example, for normally distributed data, the Blom (1958) plotting position is used with $b = 3/8$, while for data distributed according to Extreme value type 1 distribution, the Gringorten (1963) formula ($b = 0.44$) is the best For the log Pearson type 3 distribution, the optimal value for b depends on the value of the coefficient of skewness, being larger than 3/8 when the data are positively skewed and smaller than 3/8 when the data are negatively skewed However, a value of $b = 3/8$ has been used for Log Pearson type 3 distribution The same plotting positions can be applied to the logarithms of the data, when using the log normal distribution.

The cumulative distribution function is now calculated as

$$F(x) = P(X \leq x)$$
$$= 1 - p(X \geq x) \qquad (4\ 2)$$

The curve for the function is an S-curve which varies from 0 and 1 In the graph F(x) is plotted on the ordinate and the variable X on the abscissa respectively

c) K-X relationship

A plot between the frequency factor (K), as the abscissa and the variable (X), as the ordinate is also plotted. By definition, this will be a straight line The frequency factors are calculated for different distributions (Section 2 5)

d) Confidence bands

Confidence bands for 95 % confidence level are drawn to the fitted distribution. The upper and the lower confidence limits for the fitted distribution is calculated (Section 2 9) The confidence limits are represented by '/' in the plot.

e) Fitted distribution

The fitted distribution is also drawn in the plot which is represented by '&' in the plot

4.1.2 Tests for Outliers

The tests for outliers in the given data are conducted as explained in Section 2 7 When an outlier is detected, it is considered to be a rare event that has occurred as historical data This means that it is not discarded but that its plotting position computed in the normal way is considered to be

appropriate , while the value is not Hence, these rare events should be considered with care when visually evaluating the goodness of fit for the probability plots

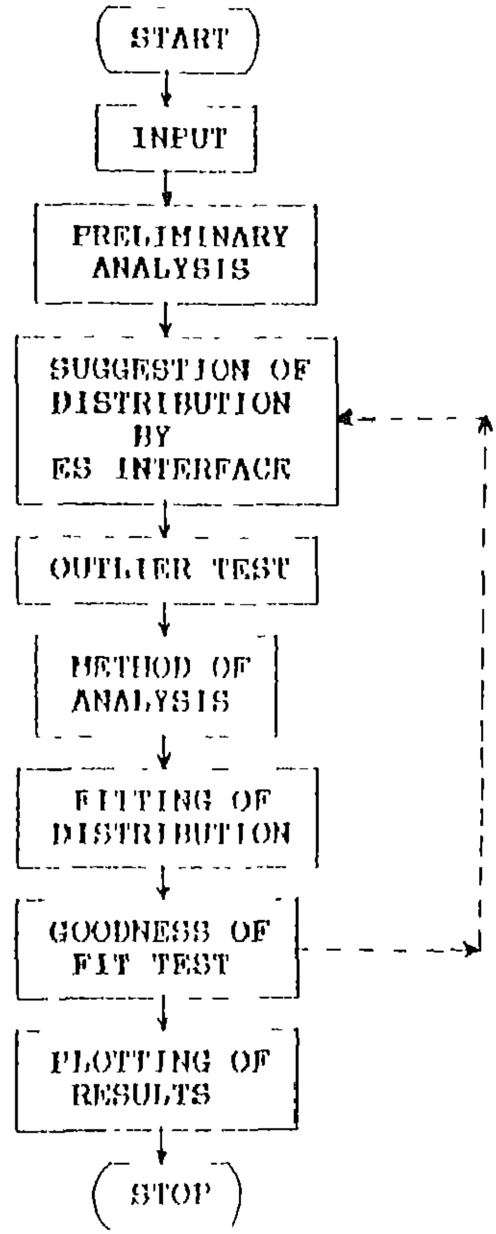
4.2 Structure

The modified version of FACHVES (Frequency Analysis of Continuous Hydrologic Variables), FACHVES Ver 2 Ø is an interactive menu based program with an embedded expert system to provide decision making support for frequency analysis of hydrologic data The program essentially consists of four parts, a main program, subroutines, ES interface and a rule based knowledge base In addition to the FACHVES program, the subroutines OUTLIER, MLS, CONFBND, FREQPL, PRPLOT , TPLOT and GOFITST have been introduced. In the main program, an additional menu OUTLIER TESTS has been added to the menu PRELIMINARY ANALYSIS

Subroutine OUTLIER tests for the presence of outlier if any and identifies the number and values of outliers present in the data for any specific distribution In case no outlier is present, it sets the number of high and low outliers to zero and these are passed on to the main program

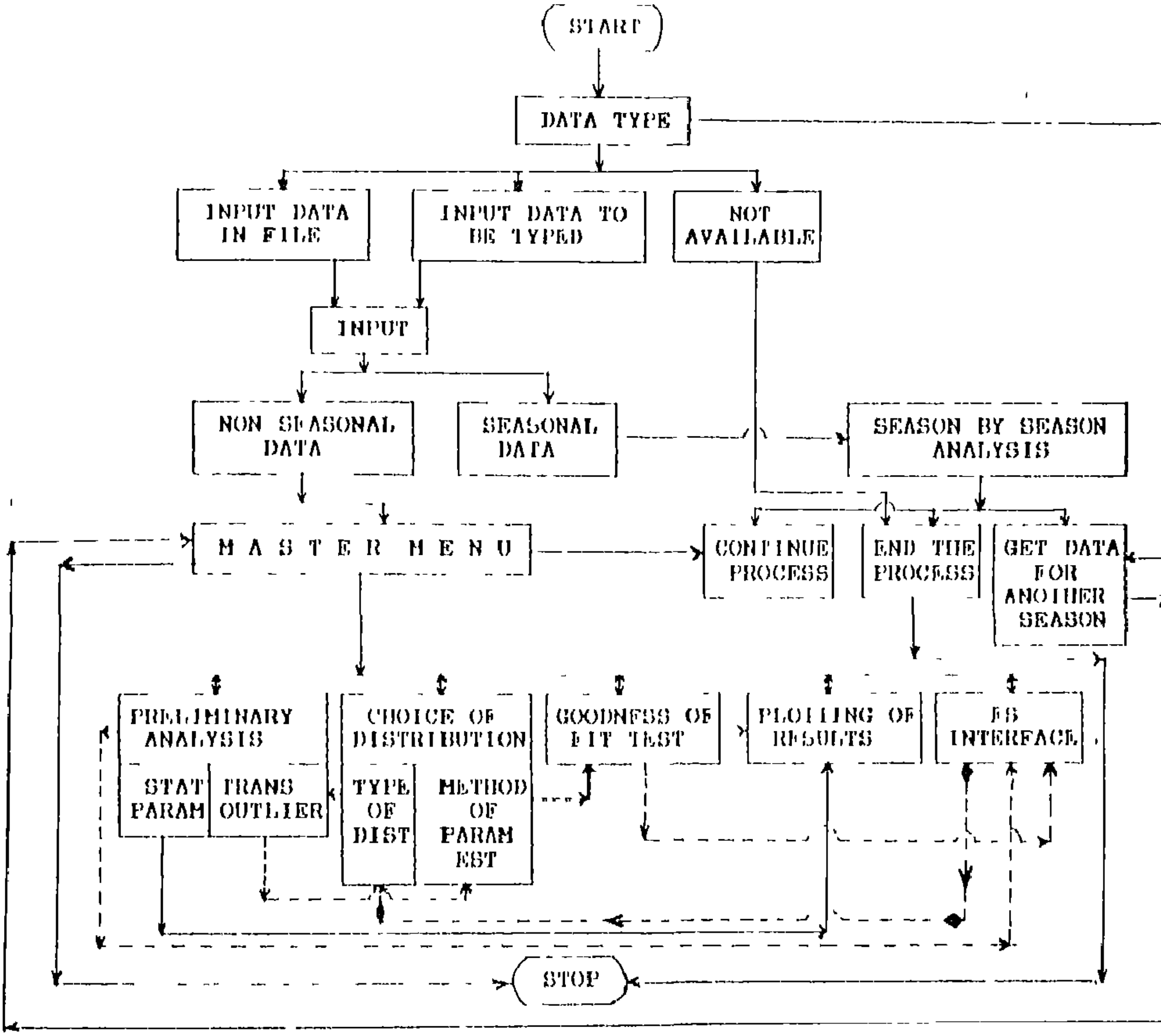
Subroutine MLS calculates the method of least squares parameter estimates for a given distribution Given a specific distribution , it estimates the plotting position by using a plotting position formula, determines the frequency factor and fits a linear regression equation between the variable and the frequency factor The Subroutine FREQPL plots the histogram for the given data, the subroutine TPLOT plots the cumulative distribution function and the Subroutine CONFBND uses a Subroutine PRPLOT to plot the fitted distribution and confidence bands in terms of the variable and the frequency factor A subroutine GOFITST compares observed and theoretical Chi square values to determine the goodness of fit of the frequency distribution

Except for the above modifications, which define Ver 2 Ø, the structure of the program is the same as given by Mishra (199Ø). The flow chart of the execution of the program, flow chart of the main program, details of the submenus and the list of subroutines are given in Figs 4 1 to 4.3 and Table 4.1



Schematic flow chart of execution of the program

Fig. 4.1



Dotted lines show the general flow chart recommended

Schematic flow chart of main program

Fig 4 2

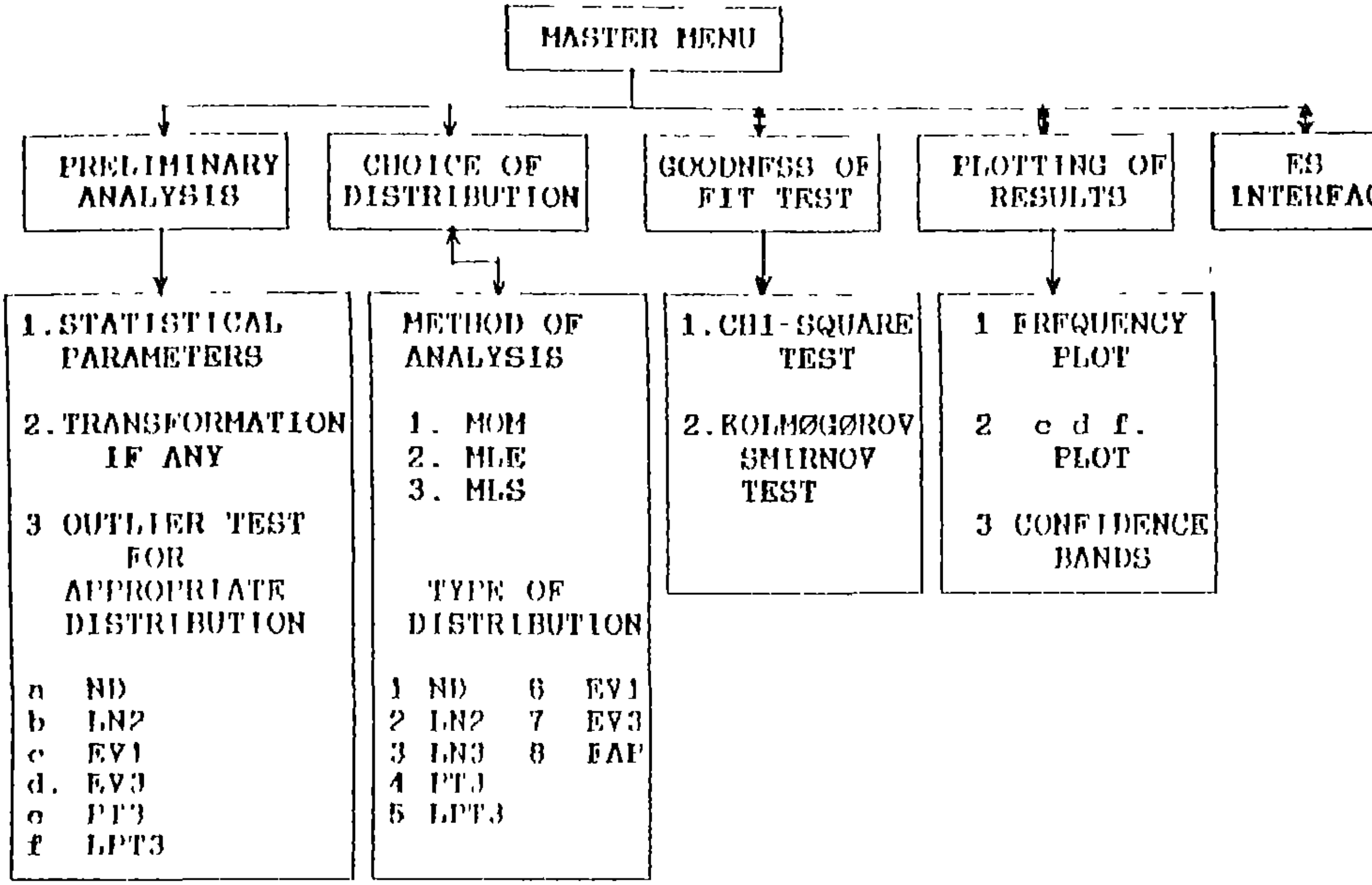


Fig 4 3 Details of Sub menus

Table 4 1

List of Main Subroutines

Subroutine	Function
ADVICE	Passes ES advice to main program
CHISQRT	Chi-square test
CONFBND	Plots the confidence bands, the fitted distribution in terms od the variable and frequency factor using the PRPLOT subroutine
FAP	Fits different distributions for seasonal data, calculates seasonal variations and tests the goodness of fit (Chi-square test)
FREQPL	Plots the histogram for the given data
GOFITST	Compares the observed and theoretical Chi-square values
IPT	Transforms the data by Inverse Pearson transformation
LN2	Fits a 2 parameter Log Normal distribution and estimates T year events
LN3	Fits a 3 parameter Log Normal distribution and estimates T year events
LOADC	Converts a C string to FORTRAN string
LOGTRAN	Transforms the data logarithmically
LP3	Fits a Log Pearson Type 3 distribution and estimates T year events
MLS	Calculates the method of least squares parameter estimates for a given distribution
ND	Fits a Normal distribution
OUTLIER	Identifies the number and the values of the outliers present in a given distribution
PARAM	Statistical parameter calculation
PRPLOT	Plotting subroutine
PTHELP	ES interface, calls ES program

PT3	Fits a Pearson Type 3 distribution and estimates T year events
SER	Computation of the standard errors of events computed from various probability distributions compared to the observed event magnitude
SQTRAN	Transforms the data by Square root transformation
STOREC	Converts a FORTRAN string to C string
T1E	Fits a Type 1 Extremal distribution and estimates T year events
T3E	Fits a Type 3 Extremal distribution and estimates T year events
TPLOT	Plots the cumulative distribution function
TSPT	Transforms the data by Two step power transformation

respectively The Subroutines MLS, OUTLIERand FREQPL are given in Appendix A

4.3 Introduction to the use of FACHVES Ver 2.0

4.3.1 Estimation of Parameters

The parameters are estimated at two stages during the execution of the expert system They are first estimated during the preliminary analysis. At this stage, parameters are estimated by the method of moments In the second stage of estimation, the choice is left to the user to choose his preference estimation of parameters This is done after a suitable distribution has been finalised after preliminary analysis and after obtaining preliminary advice from the expert system wherein all the possible distributions are indicated

The parameters can be estimated by any one of the three methods viz , the method of moments, method of least squares and method of maximum likelyhood (Section 2 5) But MLS is suggested in case outliers are detected.

4.3.2 Fitting a Distribution

After Outliers if present, have been identified and the statistics calculated, preliminary advice is taken from the expert system wherein all the suitable distributions are indicated These distributions are now fitted to the data using different methods of parameter estimation Each distribution is tested for goodness of fit, viz , by Chi square test or Kolmogorov-Smirnov test (not implemented) The distribution with the least Chi square statistic is generally assumed to fit the data best Apart from this, the goodness of fit is also observed visually by plotting the results The given data and the fitted distribution along with the confidence bands are plotted and then finally the most suitable probability distribution that fits the data is decided upon

The use of FACHVES Ver 2 Ø for fitting probability distribution to data is demonstrated with a number of examples in Chapter 5

CHAPTER 5

FREQUENCY ANALYSIS USING FACHVES Ver 2.0

5.1 Data used

The data used to test the developed program consist of one set of seasonal data and two sets of non-seasonal data (Table 5 1)

A Non seasonal data

i) Annual peak discharge for river Narmada (1951-1982)

ii) Monsoon flows for river Mahanadi at Hirakud (1946-1982)

B Seasonal data

Streamflow data for Mahanadi Basin at Hirakud dam site are used for season by season analysis The data are for the period 1931 to 1980 for 6 periods comprising of monthly data from June to October and Non monsoon data

5.2 Results and Discussions

The capabilities of the program can be tested for

a) Analysis for nonseasonal data and

b) Season by season analysis (for seasonal data)

5.2.1 Nonseasonal data

Analysis of two sets of data have been done and the interactive session with the program for one of them is reported completely in Table 5 2

i) Set 1 The analysis is done for annual peak discharge in river Narmada The program is executed by typing the name of the executable file BADDU followed by RETURN The program begins its execution and the Input module comes across first It asks whether the input will be given on the screen Since the input is given in an input file, 2 is entered The format required for the input file and other details are displayed on the screen The grammer for the input file, given also in the screen is as follows the input file should be named as TEST INP with the TITLE given in a 80A1 format followed by the number of data values, number of seasons, number of points (NCLASS + 1, where NCLASS is the number of classes into which data can be divided for Chi-square test) and then the observational data Since, such a file is available, 1 is entered The Master menu is now displayed Initially the

TABLE 5.1

Data Used in the Study

A THE ANNUAL PEAK FLOOD DISCHARGE AT MORTAKKA FOR NARMADA (1951-1982)

28,1,7,0
11127.0,13631.0,19521 0,33915 0,20746 0,11982 0,25023.0,31604 0,16135 0,
23438.0,18591.0,11338 0,19690 0,31604 0,27935.0,41691 0,18101.0,47851.0,
54063 0,36562.0,33278.0,17713 0,24354 0,29564.0,26232.0,22751.0,25662.0,
16602.0

B. HIRAKUD FLOWS(MONSON) IN THM (1946-82)

37,1,7,0

5857.00	4627 60	4552.90	4002 70	4131.40	2336 30
4491.30	3318.70	1986.30	3670.90	4116 70	3371.60
4203.70	4454.80	4029 30	9044 40	1905.50	3696 30
5773.60	1492.60	3959.10	3910.10	3570.30	3483.30
4466 30	4717.90	2301.70	5075.40	1700.50	4037 10
3134 60	3743.40	3506 40	1775 50	4471.80	2531.30
2333.40					

C. HIRAKUD FLOWS(SEASONAL) IN MCM, MAHANADI BASIN AT HIRAKUD (1931-1980)

300,6,7,0

569.0	4609.0	18067.0	8269.0	5958.0	1874.0
568.0	9129.0	12974.0	13367.0	2787.0	1941.0
2506.0	13294.0	14448.0	13277.0	2772.0	2316.0
1927.0	7588.0	17590.0	11888.0	3186.0	2109.0
856.0	10533.0	11349.0	15733.0	1273.0	1887.0
3922.0	12495.0	17656.0	10726.0	3260.0	2403.0
1272.0	10132.0	16585.0	14276.0	4172.0	2322.0
2211.0	9956.0	14606.0	13439.0	5077.0	2264.0
1603.0	7448.0	20280.0	8787.0	2151.0	2063.0
1499.0	12148.0	16329.0	7776.0	2617.0	2018.0
1709.0	7209.0	8274.0	4187.0	2043.0	1171.0
1356.0	10140.0	16742.0	18419.0	2329.0	2449.0
1121.0	10449.0	16536.0	18083.0	3199.0	2169.0
430.0	12623.0	17490.0	13474.0	4573.0	2379.0
1366.0	9633.0	13484.0	18024.0	3516.0	2301.0
2983.0	13146.0	18860.0	9207.0	3677.0	2394.0
584.0	10400.0	17632.0	15397.0	3787.0	2390.0
1385.0	7794.0	20229.0	8138.0	9837.0	2369.0
710.0	6626.0	18817.0	9678.0	5351.0	2059.0
917.0	11824.0	20666.0	7596.0	1473.0	2124.0
456.0	4805.0	24313.0	6679.0	2830.0	1454.0
789.0	8158.0	15012.0	19140.0	2976.0	2304.0
309.0	10010.0	15735.0	9512.0	2090.0	1881.0
437.0	3351.0	11224.0	1552.0	2107.0	1634.0
1600.0	7659.0	22855.0	13270.0	6571.0	2098.0
3117.0	14247.0	16360.0	13158.0	2853.0	2487.0
207.0	6038.0	12544.0	4276.0	1621.0	1234.0
449.0	13866.0	11659.0	13997.0	5997.0	2298.0
771.0	5784.0	17974.0	17649.0	4492.0	2333.0
633.0	7433.0	21939.0	5562.0	4547.0	2006.0
3691.0	26403.0	18263.0	30460.0	7200.0	4301.0
491.0	5388.0	7350.0	6117.0	1322.0	1028.0
967.0	5641.0	11341.0	14464.0	2666.0	1054.0
3134.0	15189.0	21574.0	9196.0	5101.0	2710.0
457.0	3423.0	6362.0	7920.0	1142.0	966.0
2672.0	9518.0	8716.0	6144.0	1150.0	1410.0
1326.0	6427.0	20391.0	9874.0	1732.0	1988.0
964.0	8239.0	13740.0	9976.0	2161.0	1753.0
591.0	8030.0	12380.0	7807.0	1753.0	1528.0
1768.0	12575.0	17072.0	10757.0	2425.0	2230.0
6419.0	18381.0	13358.0	13081.0	4346.0	3274.0
152.0	4761.0	8773.0	7464.0	2153.0	1167.0
252.0	10088.0	14757.0	16431.0	9278.0	2550.0
543.0	3250.0	10240.0	1673.0	1406.0	856.0
468.0	8010.0	13217.0	8924.0	4953.0	2029.0
151.0	6283.0	15742.0	8339.0	777.0	1559.0
1547.0	9231.0	13375.0	9925.0	2387.0	1821.0
1098.0	6483.0	13587.0	7474.0	1493.0	1757.0
677.0	2463.0	7440.0	1700.0	1709.0	699.0
2585.0	11571.0	9985.0	19385.0	2046.0	2379.0

Table 5.2 Interactive Session I
(Annual Peak Discharge in River Narmada)

```
**********************************************
*       EXPERT SYSTEM FOR FREQUENCY ANALYSIS  **
*            OF HYDROLOGIC DATA               **
*              BY O P MISHRA                  **
*           IMPROVED BY L K C RAO             **
*        UNDER  THE  GUIDANCE  OF             **
*           Dr S Ramaseshan                   **
*           at I I T Kanpur                   **
**********************************************
*    WISH   YOU    HAPPY COMPUTING            **
*             W I T H                         **
*        F A C H V E S Ver 3 0                **
**********************************************
* *--* FREQUENCY ANALYSIS WITH EXPERT ADVICE *--*
*  -  -- -   -               ----    -- - --- *
***************************************
*     I N P U T    M O D U L E        *
* -----  -- ----------- ---------- -----*
Do you want to give input on the screen?
If yes, Type 1 ,If no, Type 2
    If you want to quit Type 3
2
--The input file should be named as TEST.INP
--         It must have following
--       *    a TITLE with format 80A1
--              AND, unformatted
-        *    NX--Number of data
--       *    NS -Number of seasons
--         (for annual series NS  1)
--       *    NPT -Number of points
--            NPT=NCLASS+1 where NCLASS is number of
                classes in which data can be divided
                for chisquare test  The class frequency
                should be equal to 5
--              (It can be 7/11/13)
--       *    Observational hydrologic data RX(I)
--       (Not more than 600;
                 seasons(in row)/years(in column)
Have you got such an input file ? If yes,type 1
       If you want to quit ,type 2
       If you want to go to previous menu type 3
1
*------ --- -- -------------- --------------*
*       MASTER-MENU                         *
*------- -----   ----    -- - --------------*
*   1 PRELIMINARY ANALYSIS                  *
*   2 CHOICE OF DISTRIBUTIONS               *
*   3 TESTS FOR GOODNESS OF FIT             *
*   4 PLOTTING OF RESULTS                   *
*   5 QUIT                                  *
*   6.HELP FROM EXPERT SYSTEM INTERFACE     *
*-------  - -   -----  - --- ------ -  ------ -*
Type the serial number of desired option
1
***************************************
*   $$$ PRELIMINARY ANALYSIS $$$       *
*-----     -       --       - --- ---------*
*   1 STATISTICAL PARAMETERS           *
*   2 FREQUENCY PLOT                   *
*   3 C D F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                 *
*   5.TRANSFORMATIONS                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                   *
*   8.OUTLIER TESTS                    *
*---------  ------- -- --  ---------- - -*
Type the serial number of desired option
```

```
1
 MEAN=  25382 7852
 VAR-   0 11461E+09
 SKEW    0 95990E+00
 STDEV    0 10706E+05
 KURTOSIS-    0 40343E+01  AVG DEVIATION- 0 82410E+04

THIS ANALYSIS IS FOR LOG-TRANSFORMED DATA
 MEAN=      10 0593
 VAR-   0 17721E+00
 SKEW-    0.96092E-02
 STDEV=    0 41498E+00
 KURTOSIS=    0 28240F+01
XMIN= 0 93171E+01  XMAX= 0.10898E+02

****************************************
*   *** PRELIMINARY ANALYSIS ***         *
*---- -          - - --    -------------*
*   1 STATISTICAL PARAMETERS             *
*   2 FREQUENCY PLOT                     *
*   3 C D F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                   *
*   5 TRANSFORMATIONS                    *
*   6 HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                     *
*   8 OUTLIER TESTS                      *
*---- --  - -------- ------- -----------*
Type the serial number of desired option
6
*******************************************
*  ** EXPERT SYSTEM INTERFACE **        **
*                          -  - - ------- *
*   1 FOR PRELIMINARY ADVICE             *
*   2 FOR SECONDARY   ADVICE           ' *
*   3 FOR EXPERT ANALYSIS OF RESULTS     *
*   7 TO MASTER MENU                     *
*---- ---------------------------------*
Which help do you want?
Type the serial number of desired option
1
f-0     (initial-fact)
f-1     (start)
f-0     (initial-fact)
f-1     (start)
f-2     (mean 25382)
f-3     (skew 0 95990002)
f-4     (kurt 4 03429985)
f-5     (stdev 10706)
f-6     (stdev of logtran data 0.41497999)
f-7     (length of data 28)
       W E L C O M E

           TO

         F A C H U R S

This  expert   system   is  designed  for  single
               peaked  continuous hydrologic variables   only

IF YOU DO NOT UNDERSTANT ANY QUESTION ; TYPE nk
              nk --- Not known

Type of data       type the first letter (e/a/s)
               Whether it is exceedence OR annual "s
Do you think that outliers are present?
```

nk

```
In absence of information outliers are
              initially assumed not to be present

Use method of moments for parameter estimation

Here is the list of steps for analysis      **
              1 **DO THE PRELIMINARY ANALYSIS
              2 **TEST FOR OUTLIERS
              3 **FIT A SUITABLE DISTRIBUTION
              4 **TEST FOR GOODNESS OF FIT
              5 **PLOT THE RESULTS
You can fit
              1 a pearson type 3 distribution
              2 a lognormal 3 parameter distribution
              3 a logpearson type 3 distribution
If pearson type 3 is not suggested then
              Gamma(2 parameters) may be fit
1 You can fit either one OR all available distributions

 ** If more than one distribution satisfies the
                goodness of fit test (chi square test) then
                the best fit is the one with the least
                Chi square value

PROCESS IS BEING STOPPED;
              RESET THE PROCESS IF YOU WANT TO CONTINUE
              WITH OTHER SET OF DATA
Do you want any help ?  no
Suggestions passed should be taken in sequencial order
f 0      (initial-fact)
f-1      (start)
f-2      (mean 25387)
f-3      (skew 0 75990002)
f-4      (kurt 4 03429985)
f-5      (stdev 10706)
f-6      (stdev of logtran data 0 41497999)
f-7      (length of data 38)
f-8      (type of data a)
f-9      (process continue yes)
f-10     (data is large enough)
f-11     (outliers are nk)
f-12     (outliers no)
f-13     (positive skew)
f-14     (logtran val is 2 49346423)
f-15     (suggestion 15)
f-16     (esad PT3,LP3,LN3)
f-17     (suggestion 10)
f-18     (esad GD)
f 19     (stop)
f 20     (help no)
Excising rule  rule1
Excising rule' rule2
Excising rule  rule3
Excising rule  rule4
Excising rule  rule5
Excising rule  rule6
Excising rule' rule7
Excising rule  rule8
Excising rule  rule9
Excising rule' rule10
Excising rule  rule11
Excising rule  rule12
Excising rule  rule13
Excising rule  rule14
Excising rule' rule15
Excising rule' rule16
Excising rule  rule18
Excising rule  rule19
```

```
Excising rule: rule21
Excising rule: rule22
Excising rule: rule23
Excising rule: rule24
Excising rule: rule25
Excising rule: rule26
Now you are out of clips
Value passed =     15 00000
Name passed is =
PT3,LP3,LN3
*-------------------------------------------*
*       MASTER-MENU                         *
*                -   ---------------------- *
*   1 PRELIMINARY ANALYSIS                  *
*   2 CHOICE OF DISTRIBUTIONS               *
*   3 TESTS FOR GOODNESS OF FIT             *
*   4 PLOTTING OF RESULTS                   *
*   5 QUIT                                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE     *
*-------------------------------------------*
VALUE PASSED IS     15 00000
NAME PASSED IS
PT3,LP3,LN3
Type the serial number of desired option
1
****************************************
*   $$$ PRELIMINARY ANALYSIS $$$         *
*----------------------------------------*
*   1.STATISTICAL PARAMETERS             *
*   2.FREQUENCY PLOT                     *
*   3 C D.F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                   *
*   5 TRANSFORMATIONS                    *
*   6.HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                     *
*   8 OUTLIER TESTS                      *
*----------------------------------------*
Type the serial number of desired option
8
 MEAN=  25382 2852
 VAR=   0 11461E+09
 SKEW=   0.95990E+00
 STDEV=   0.10706E+05
 KURTOSIS=   0 40343E+01  AVG DEVIATION= 0 83410E+04

 THE DATA IS BEING CHECKED FOR OUTLIERS
 FOR WHICH DISTRIBUTION DO YOU WANT IT CHECKED i e
 WHETHER 1. NORMAL,  2. LOG-NORMAL(2 PARAMETERS)
 3. LOG-NORMAL(3 PARAMETERS),  4. EV-I
 5  EV-III,  6  PT-III,  7  LPT-III
```

```
6MEAN=  25387 2852
 VAR=   0 11461E+09
 SKEW=   0.95990E+00
 STDEV=   0 10706E+05
 KURTOSIS=   0 40343E+01  AVG DEVIATION= 0 82410E+04

 THE PARAMETERS FOR THE ORIGINAL DATA ARE
 MEAN=  25387 2852
 SKEW=   0.95990E+00
 STDEV=   0.10706E+05

 MEAN=       0 4441
 VAR=   0 31540E+00
 SKEW=   0 18020E+01
 STDEV=   0 56160E+00
 KURTOSIS=   0 58336E+01  AVG DEVIATION= 0 43109E+00
 *** FOR ORIGINAL DATA ***
CHECK FOR OUTLIERS BY WRC METHOD

 THE PARAMETERS FOR THE LOG-TRANSFORMED DATA ARE :
 MEAN=       0 0000
 SKEW=   0 00000E+00
 STDEV=   0 00000E+00

 XMIN=  0 67419E 01      XMAX=  0 21432E+01
 FOR NORMAL DISTRIBUTION
 THE THRESHOLD VALUE FOR HIGHER OUTLIERS :    1.867318
 THE THRESHOLD VALUE FOR LOWER OUTLIERS :  -0 9789862
XMIN= 0 67419E-01  XMAX= 0 21432E+01
NO OF HIGH OUTLIERS =      1
NO OF LOW OUTLIERS=      0
 THE NEW PARAMETERS OBTAINED DUE TO THE
      PRESENCE OF OUTLIERS ARE :
 MEAN=  24320 0371
 VAR=   0 86213E+08
 SKEW=   0 68045E+00
 STDEV=   0 92851E+04
 KURTOSIS=   0 34592E+01  AVG DEVIATION= 0.73418E+04
*************************************
*   $$$ PRELIMINARY ANALYSIS $$$        *
*---------- -  -           - ---------   --*
*   1 STATISTICAL PARAMETERS               *
*   2 FREQUENCY PLOT                       *
*   3 C D.F.PLOT (FOR ALL DISTRIBUTIONS)  *
*   4.CONFIDENCE BANDS                     *
*   5 TRANSFORMATIONS                      *
*   6 HELP FROM EXPERT SYSTEM INTERFACE   *
*   7 TO MASTER MENU                       *
*   8 OUTLIER TESTS                        *
*---------------------------------------*
Type the serial number of desired option
6
****************************************
*  $$ EXPERT SYSTEM INTERFACE $$       **
*---  -               -  -- ---- ---*
*   1 FOR PRELIMINARY ADVICE            *
*   2 FOR SECONDARY  ADVICE             *
*   3 FOR EXPERT ANALYSIS OF RESULTS    *
*   7 TO MASTER MENU                    *
* - --     -          --- - -----------*
Which help do you want?
Type the serial number of desired option
```

```
?
f-0     (initial-fact)
f-1     (start)
f 0     (initial fact)
f 1     (start)
f 2     (mean 25382)
f 3     (skew 0 95990002)
f 4     (kurt 4 03429985)
f 5     (stdev 10706)
f-6     (stdev of logtran data 0 41497999)
f-7     (length of data 28)
      W E L C O M E

         TO

       F A C H V E S

This  e pert   system   is  designed  for  single
               peaked  continuous hydrologic variables   only

IF YOU DO NOT UNDERSTANT ANY QUESTION , TYPE nk
              nk --- Not known

Type of data ---- type the first letter (e/a/s)
               Whether it is exceedence OR annual ?a
Do you think that outliers are present?  yes
You should use method of least squares for
                parameter estimation
You can fit
               1 a pearson type 3 distribution
               2 a lognormal 3 parameter distribution
               3 a logpearson type 3 distribution
If pearson type 3 is not suggested then
               Gamma(2 parameters) may be fit
I You can fit either one OR all available distributions

 ** If more than one distribution satisfies the
                 goodness of fit test (chi square test) then
                 the best fit is the one with the least
                 chi square value

PROCESS IS BEING STOPPED;
               RESET THE PROCESS IF YOU WANT TO CONTINUE
               WITH OTHER SET OF DATA
Do you want any help ?  no
Suggestions passed should be taken in sequencial order
f-0     (initial-fact)
f-1     (start)
f-2     (mean 25382)
f-3     (skew 0 95990002)
f-4     (kurt 4 03429985)
f 5     (stdev 10706)
f-6     (stdev of logtran data 0.41497999)
f-7     (length of data 28)
f-8     (type of data a)
f-9     (process continue yes)
f 10    (data is large enough)
f 11    (outliers are yes)
f-12    (outliers yes)
f-13    (outliers present)
f-14    (positive skew)
f-15    (logtran val is 3 49346473)
f-16    (suggestion 15)
f-17    (esad IT3,IT3,IH3)
f-18    (suggestion 10)
f 19    (read GP)
f 20    (stop)
f 21    (help no)
Excising rule  rule1
Excising rule  rule2
Excising rule: rule3
```

```
F cising rule  rule4
F cising rule  rule5
F cising rule  rule6
F rising rule' rule7
F rising rule  rule8
F rising rule  rule9
F rising rule  rule10
F rising rule  rule11
F rising rule  rule12
F rising rule  rule13
F cising rule  rule14
F rising rule' rule15
F rising rule  rule16
F rising rule  rule18
F rising rule  rule19
E cising rule' rule20
F cising rule  rule21
F rising rule' rule22
F rising rule' rule23
F cising rule  rule24
F cising rule' rule25
F rising rule  rule26
Now you are out of clips
Value passed -    15 00000
Name passed is =
FT3,LP3,LN3
*-----------------------------------------*
*      MASTER MENU                        *
*                 ----------------------- *
*   1 PRELIMINARY ANALYSIS                *
*   2 CHOICE OF DISTRIBUTIONS             *
*   3 TESTS FOR GOODNESS OF FIT           *
*   4 PLOTTING OF RESULTS                 *
*   5 QUIT                                *
*   6 HELP FROM EXPERT SYSTEM INTERFACE   *
*-----------------------------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
FT3,LT3,LN3
Type the serial number of desired option
2
***************************************
*  $$ METHOD OF ANALYSIS $$          *
* -- -- - --- --   ------------ --   *
*   1 METHOD OF MOMENTS              *
*   2 METHOD OF MAXIMUM LIKELYHOOD   *
*   3.METHOD OF LEAST-SQUARES        *
*   4 TO MASTER MENU                 *
*------------------------------------*
Which method do you like?
Type the serial number of desired option
3
It is only available for Normal Distribution
For other cases the data is transformed into normal  variables and then MLS is u
sed
********************************
**   $$ CHOICE OF DISTRIBUTION $$      **
*-        - ---- ----------------------*
*   1.NORMAL                  *
*   2 LOG NORMAL (2 PARAMETER)  *
*   3 LOG NORMAL (3 PARAMETER)  *
*   4 PEARSON TYPE 3            *
*   5 LOG PEARSON TYPE 3        *
*   6 EXTREME VALUE TYPE 1      *
*   7 EXTREME VALUE TYPE-3      *
*   8 EAP--PROGRAMS             *
*   9 TO MASTER MENU            *
*  10 TRANSFORMATIONS           *
*  11 METHOD OF ANALYSIS        *
*--   - - ---------  -------------------*
If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number
```

PEARSON AND LOG PEARSON DISTRIBUTION
****

I	Y(I)	LX(I)	P(I)	KTP(I)	KTLP(I)
2	47851.00	4.68	0.06	1.76	1.58
3	41691.00	4.62	0.09	1.39	1.37
4	36561.00	4.56	0.13	1.13	1.13
5	33915.00	4.53	0.16	0.93	0.98
6	33278.00	4.52	0.20	0.76	0.84
7	31604.00	4.50	0.23	0.62	0.72
8	31604.00	4.50	0.27	0.49	0.61
9	29564.00	4.47	0.31	0.37	0.51
10	27935.00	4.45	0.34	0.26	0.41
11	26233.00	4.42	0.38	0.16	0.31
12	25662.00	4.41	0.41	0.06	0.22
13	25023.00	4.40	0.45	-0.03	0.13
14	24354.00	4.39	0.48	-0.11	0.04
15	23438.00	4.37	0.52	-0.20	-0.05
16	22751.00	4.36	0.55	-0.28	-0.13
17	20746.00	4.32	0.59	-0.36	-0.23
18	19690.00	4.29	0.62	-0.44	-0.32
19	19521.00	4.29	0.66	-0.52	-0.41
20	18591.00	4.27	0.69	-0.60	-0.51
21	18101.00	4.26	0.73	-0.68	-0.61
22	17713.00	4.25	0.77	-0.76	-0.72
23	16602.00	4.22	0.80	-0.85	-0.85
24	16135.00	4.21	0.84	-0.94	-0.98
25	13631.00	4.13	0.87	-1.04	-1.13
26	11982.00	4.08	0.91	-1.16	-1.32
27	11338.00	4.05	0.94	-1.30	-1.57
28	11127.00	4.05	0.98	1.50	-2.01

THE CHI SQUARE STATISTIC OBTAINED DUE TO THE
PRESENCE OF OUTLIERS USING MLS IS :

CHI SQUARE STATISTIC
****

1.307353

NO OF DEGREES OF FREEDOM
****

3

THE THEORITICAL CHI-SQUARE VALUE IS : 7.815000

FOR 3 DEGREES OF FREEDOM

THE FIT IS GOOD

PEARSON TYPE 3 DISTRIBUTION
****

METHOD OF LEAST SQUARES
------ -- ----- -------

ALPHA =	4456.367	MEAN =	24320.04
BETA =	4.341196	STDEV =	9285.079
GAMMA =	4974.078	SKEW =	0.9598985

```
*********************************************
*  ** TEST OF GOODNESS OF FIT **           **
*-------------------------------------------*
*     1 CHI-SQUARE TEST                     *
*     2.KOLMOGOROV SMIRNOV TEST             *
*     3 TO PLOTTING RESULTS                 *
*     4 COMPARISION OF STANDARD ERROR       *
*       OF ESTIMATE FOR DIFFERENT           *
*         DISTRIBUTIONS                     *
*     5 TO MASTER MENU                      *
*     6 HELP FROM EXPERT-SYSTEM INTERFACE   *
*-------------------------------------------*
Type the serial number of desired option
1
 THE NEW PARAMETERS OBTAINED DUE TO THE
       PRESENCE OF OUTLIERS ARE
 MEAN=  34370 0371
 VAR-   0 86713E+08
 SKEW-   0 68045E+00
 STDEV-   0 97851E+04
 KURTOSIS=   0 34597E+01  AVG DEVIATION= 0 73418E+04

 THE CHI SQUARE STATISTIC OBTAINED DUE TO
 THE PRESENCE OF OUTLIERS USING MM IS

CHI SQUARE STATISIIC
********************

  1 307353

NO  OF DEGREES OF FREEDOM
*************************

          3

 THE THEORITICAL CHI-SQUARE VALUE IS •    7 815000

 FOR              3   DEGREES OF FREEDOM

 THE FIT IS GOOD

 THE CHI SQUARE STATISTIC OBTAINED WITHOUT
 THE PRESENCE OF OUTLIERS USING MM IS

CHI-SQUARE STATISTIC
********************

  2.109704

NO  OF DEGREES OF FREEDOM
*************************

          3

 THE THEORITICAL CHI-SQUARE VALUE IS      7 815000

 FOR              3   DEGREES OF FREEDOM

 THE FIT IS GOOD
*********************************************
**  ** PLOTTING RESULTS                    *
*-------------------------------------------*
*   1 FREQUENCY PLOT                        *
*   2 C D F PLOT                            *
*   3 CONFIDENCE BANDS                      *
*   4 TO MASTER MENU                        *
*-------------------------------------------*
Type the serial number of desired option
```

```
4
*-----------------------------------------------*
*      MASTER MENU                              *
*-----------------------------------------------*
*   1 PRELIMINARY ANALYSIS                      *
*   2 CHOICE OF DISTRIBUTIONS                   *
*   3 TESTS FOR GOODNESS OF FIT                 *
*   4 PLOTTING OF RESULTS                       *
*   5 QUIT                                      *
*   6 HELP FROM EXPERT SYSTEM INTERFACE         *
*-----------------------------------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
PT3,PT3,LN3
Type the serial number of desired option
1
****************************************
*   *** PRELIMINARY ANALYSIS ***       *
*--------------------------------------*
*   1 STATISTICAL PARAMETERS           *
*   2 FREQUENCY PLOT                   *
*   3 C D F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                 *
*   5 TRANSFORMATIONS                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                   *
*   8 OUTLIER TESTS                    *
*--------------------------------------*
Type the serial number of desired option
8
 MEAN=  25382.2852
 VAR=   0 11161E+09
 SKEW=   0 95990E+00
 STDEV=   0 10706E+05
 KURTOSIS=   0 40343E+01  AVG DEVIATION= 0 82410E+04

 THE DATA IS BEING CHECKED FOR OUTLIERS
 FOR WHICH DISTRIBUTION DO YOU WANT IT CHECKED i e
 WHETHER 1  NORMAL,  2  LOG-NORMAL(2 PARAMETERS)
 3  LOG-NORMAL(3 PARAMETERS),  4  EV-I
 5  EV-III,  6  PT III,  7  LPT-III
3
 MEAN=      10 0593
 VAR=   0 17221E+00
 SKEW=   0 96093E 02
 STDEV=   0 41498E+00
 KURTOSIS=   0 28240E+01  AVG DEVIATION= 0 33178E+00
 *** FOR ORIGINAL DATA ***
CHECK FOR OUTLIERS BY WRC METHOD

 THE PARAMETERS FOR THE LOG-TRANSFORMED DATA ARE :
 MEAN=      10 0593
 SKEW=   0 96093E 02
 STDEV=   0 41498E+00

 XMIN=  0 93171E+01      XMAX=  0 10898E+02
 FOR LOG NORMAL DISTRIBUTION
 THE THRESHOLD VALUE FOR HIGHER OUTLIERS :     66894.96
 THE THRESHOLD VALUE FOR LOWER OUTLIERS :    8166 022
XMIN= 0 11122E+05  XMAX= 0 54063E+05
NO OF HIGH OUTLIERS =      0
NO OF LOW OUTLIERS=      0
 OUTLIERS ARE ABSENT FOR THIS DISTRIBUTION
```

```
*****************************************
*   *** PRELIMINARY ANALYSIS ***        *
* -  --- -------   --- --------------- *
*   1 STATISTICAL PARAMETERS            *
*   2 FREQUENCY PLOT                    *
*   3 C.D.F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                  *
*   5 TRANSFORMATIONS                   *
*   6 HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                    *
*   8 OUTLIER TESTS                     *
*- ------------- -  -- -- ------- -----*
Type the serial number of desired option
7
*-------   -- ---- -------------------------*
*      MASTER-MENU                          *
*----------- ------ -------------------------*
*   1 PRELIMINARY ANALYSIS                  *
*   2 CHOICE OF DISTRIBUTIONS               *
*   3 TESTS FOR GOODNESS OF FIT             *
*   4 PLOTTING OF RESULTS                   *
*   5 QUIT                                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE     *
* ----- -  ------- ------ ------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
PT3,LP3,LN3
Type the serial number of desired option
2
*****************************************
*  ** METHOD OF ANALYSIS **         *
*-----------------------------------*
*   1 METHOD OF MOMENTS             *
*   2 METHOD OF MAXIMUM LIKELYHOOD   *
*   3.METHOD OF LEAST-SQUARES        *
*   4 TO MASTER MENU                *
* ----------------------------------*
Which method do you like?
Type the serial number of desired option
1
*************************************
**   ** CHOICE OF DISTRIBUTION **     **
*----    -   ----------  ----- --------*
*   1 NORMAL                    *
*   2 LOG-NORMAL (2 PARAMETER)  *
*   3 LOG NORMAL (3 PARAMETER)  *
*   4 PEARSON TYPE 3            *
*   5 LOG-PEARSON TYPE 3        *
*   6 EXTREME VALUE TYPE-1      *
*   7 EXTREME VALUE TYPE-3      *
*   8 EAP  PROGRAMS             *
*   9 TO MASTER MENU            *
*  10 TRANSFORMATIONS           *
*  11 METHOD OF ANALYSIS        *
*--- --- ---------- - ----------------*
If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number

3
******************************************
*  ** TEST OF GOODNESS OF FIT **            **
*-----    --  --- -  ------------ --- ---*
*    1 CHI SQUARE TEST                    *
*    2 KOLMOGOROV-SMIRNOV TEST            *
*    3 TO PLOTTING RESULTS                *
*    4 COMPARISION OF STANDARD ERROR      *
*      OF ESTIMATE FOR DIFFERENT          *
*        DISTRIBUTIONS                    *
*    5 TO MASTER MENU                     *
*    6 HELP FROM EXPERT-SYSTEM INTERFACE  *
*- --- - -- ---- -    -------------------*
Type the serial number of desired option
```

```
1

CHI SQUARE STATISTIC
********************

  5 665852

NO  OF DEGREES OF FREEDOM
**************************

          3

 THE THEORITICAL CHI SQUARE VALUE IS      7 815000

 FOR            3   DEGREES OF FREEDOM

 THE FIT IS GOOD
*************************************************
**  $$ PLOTTING RESULTS                       *
*-----------------------------------------------*
*   1 FREQUENCY PLOT                          *
*   2 C D F PLOT                              *
*   3 CONFIDENCE BANDS                        *
*   4 TO MASTER MENU                          *
*- ---------- ----------------- --- ---*
Type the serial number of desired option
4
*---------- - - -----------------------------*
*      MASTER-MENU                                 *
*------ ------------------------------------------*
*   1 PRELIMINARY ANALYSIS                         *
*   2 CHOICE OF DISTRIBUTIONS                      *
*   3.TESTS FOR GOODNESS OF FIT                    *
*   4 PLOTTING OF RESULTS                          *
*   5 QUIT                                         *
*   6 HELP FROM EXPERT SYSTEM INTERFACE            *
*------- - -------------------------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
FI3,LF3,LN3
Type the serial number of desired option
1
*****************************************
*   $$$ PRELIMINARY ANALYSIS $$$          *
*-----------------------------------------*
*   1 STATISTICAL PARAMETERS              *
*   2 FREQUENCY PLOT                      *
*   3 C D F.PLOT (FOR ALL DISTRIBUTIONS)  *
*   4 CONFIDENCE BANDS                    *
*   5.TRANSFORMATIONS                     *
*   6 HELP FROM EXPERT SYSTEM INTERFACE   *
*   7 TO MASTER MENU                      *
*   8 OUTLIER TESTS                       *
* --- ----- - -- - ----------- ---------*
Type the serial number of desired option
8
 MEAN=  25382 3852
 VAR-    0 11461E+09
 SKEW-    0.95990E+00
 STDEV=    0 10706E+05
 KURTOSIS=   0 40343E+01  AVG DEVIATION= 0 82410E+04

 THE DATA IS BEING CHECKED FOR OUTLIERS
 FOR WHICH DISTRIBUTION DO YOU WANT IT CHECKED i e
 WHETHER 1  NORMAL,  2  LOG-NORMAL(2 PARAMETERS)
 3  LOG NORMAL(3 PARAMETERS),  4  EV-I
 5. EV-III,  6  PT III,  7  LPT-III
```

```
7
 MEAN=      10 0593
 VAR=   0 17221E+00
 SKEW=   0 96093E-02
 STDEV=   0 41498E+00
 KURTOSIS=   0 28240E+01  AVG DEVIATION= 0 33178E+00
 MEAN=       0 4463
 VAR=   0 29468E+00
 SKEW=   0 16342E+01
 STDEV=   0 54384E+00
 KURTOSIS=   0 51557E+01  AVG DEVIATION= 0 42618E+00
 *** FOR ORIGINAL DATA ***
CHECK FOR OUTLIERS BY WRC METHOD

 THE PARAMETERS FOR THE LOG-TRANSFORMED DATA ARE
 MEAN=      10 0593
 SKEW=   0 96093E-02
 STDEV=   0 41498E+00

 XMIN=  0 83750E-02     XMAX=  0 20159E+01
 FOR NORMAL DISTRIBUTION
 THE THRESHOLD VALUE FOR HIGHER OUTLIERS       1 821831
 THE THRESHOLD VALUE FOR LOWER OUTLIERS     -0 9293031
XMIN= 0 83750E-02  XMAX= 0 20159E+01
NO OF HIGH OUTLIERS =      1
NO OF LOW OUTLIERS=      0
 THE NEW PARAMETERS OBTAINED DUE TO THE
      PRESENCE OF OUTLIERS ARE
 MEAN=  24320 0391
 VAR=   0 86213E+08
 SKEW=   0 68045E+00
 STDEV=   0 92851E+04
 KURTOSIS=   0 34592E+01  AVG DEVIATION= 0 73418E+04
*****************************************
*   *** PRELIMINARY ANALYSIS ***         *
*---------------------------------------*
*   1 STATISTICAL PARAMETERS             *
*   2 FREQUENCY PLOT                     *
*   3 C D F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                   *
*   5 TRANSFORMATIONS                    *
*   6 HELP FROM EXPERT SYSTEM INTERFACE  *
*   7 TO MASTER MENU                     *
*   8 OUTLIER TESTS                      *
*---------------------------------------*
Type the serial number of desired option
7
*------ ---     - -----------------------------*
*      MASTER-MENU                             *
*----------------------------------------------*
*   1 PRELIMINARY ANALYSIS                     *
*   2 CHOICE OF DISTRIBUTIONS                  *
*   3 TESTS FOR GOODNESS OF FIT                *
*   4 PLOTTING OF RESULTS                      *
*   5 QUIT                                     *
*   6 HELP FROM EXPERT SYSTEM INTERFACE        *
*----------------------------------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
PT3,LP3,LN3
Type the serial number of desired option
2
*******************************************
*  ** METHOD OF ANALYSIS **          *
*-----------------------------------*
*   1 METHOD OF MOMENTS              *
*   2 METHOD OF MAXIMUM LIKELYHOOD   *
*   3 METHOD OF LEAST-SQUARES        *
*   4 TO MASTER MENU                 *
*-----------------------------------*
```

[illegible]
Type the serial number of desired option
3
It is only available for Normal Distribution
For other cases the data is transformed into normal variables and then MLS is used

```
************************************
**   $$ CHOICE OF DISTRIBUTION $$      **
*------------------------------------*
*   1 NORMAL                        *
*   2 LOG-NORMAL (2-PARAMETER)      *
*   3 LOG-NORMAL (3-PARAMETER)      *
*   4 PEARSON TYPE 3                *
*   5 LOG-PEARSON TYPE-3            *
*   6 EXTREME-VALUE TYPE-1          *
*   7 EXTREME-VALUE TYPE-3          *
*   8 FAP--PROGRAMS                 *
*   9 TO MASTER MENU                *
*  10 TRANSFORMATIONS               *
*  11 METHOD OF ANALYSIS            *
*------------------------------------*
```

If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number
5

PEARSON AND LOG-PEARSON DISTRIBUTION
****

I	X(I)	LX(I)	P(I)	KTP(I)	KTLP(I)
2	47851 00	4 68	0 06	1 76	1 58
3	41691 00	4 62	0 09	1 39	1 32
4	36562 00	4 56	0 13	1 13	1 13
5	33915 00	4 53	0 16	0 93	0 98
6	33278 00	4 52	0 20	0 76	0 84
7	31604 00	4 50	0 23	0 62	0.72
8	31604 00	4 50	0 27	0 49	0 61
9	29564 00	4 47	0 31	0 37	0 51
10	27935 00	4 45	0 34	0 26	0 41
11	26232 00	4 42	0 38	0 16	0 31
12	25662 00	4 41	0 41	0 06	0 22
13	25023 00	4 40	0 45	-0 03	0 13
14	24354 00	4 39	0 48	-0 11	0 04
15	23438 00	4 37	0 52	-0 20	-0 05
16	22751 00	4 36	0 55	-0 28	-0 13
17	20746 00	4 32	0 59	-0 36	-0 23
18	19690 00	4 29	0 62	-0 44	-0 32
19	19521 00	4 29	0 66	-0 52	-0 41
20	18591 00	4 27	0 69	-0 60	-0 51
21	18101 00	4 26	0 73	-0 68	-0 61
22	17713 00	4 25	0 77	-0 76	-0 72
23	16602 00	4 22	0 80	-0 85	0 85
24	16135 00	4 21	0 84	-0 94	0 98
25	13631 00	4 13	0 87	-1 04	-1 13
26	11982 00	4 08	0 91	-1 16	-1 32
27	11338 00	4 05	0 94	-1 30	-1 57
28	11127 00	4 05	0 98	-1 50	-2 01

THE CHI SQUARE STATISTIC OBTAINED DUE TO THE
PRESENCE OF OUTLIERS USING MLS IS

CHI-SQUARE STATISTIC
****

0 3842761

NO OF DEGREES OF FREEDOM
****

3

THE THEORITICAL CHI-SQUARE VALUE IS 7 815000

FOR 3 DEGREES OF FREEDOM

THE FIT IS GOOD

```
LOG-PEARSON TYPE 3 DISTRIBUTION
*******************************

 METHOD OF LEAST SQUARES
 ----------------------

 ----------------------------------------

ALPHA =   8 1110437E-04    MEAN  =    4 355213
BETA  =    43229 07         STDEV =   0 1686416
GAMMA =   -30.70808        SKEW  =   9 6192677E-03


************************************************
*  $$ TEST OF GOODNESS OF FIT $$             **
*-------------------------------------------*
*    1 CHI-SQUARE TEST                        *
*    2 KOLMOGOROV-SMIRNOV TEST                *
*    3 TO FLOTTING RESULTS                    *
*    4 COMPARISION OF STANDARD ERROR          *
*      OF ESTIMATE FOR DIFFERENT              *
*        DISTRIBUTIONS                        *
*    5 TO MASTER MENU                         *
*    6 HELP FROM EXPERT-SYSTEM INTERFACE      *
*-------------------------------------------*
Type the serial number of desired option
1
 THE NEW PARAMETERS OBTAINED DUE TO THE
      PRESENCE OF OUTLIERS ARE
 MEAN=  24320 0371
 VAR=   0 86213E+08
 SKEW=   0 68045E+00
 STDEV=   0 92851E+04
 KURTOSIS=   0 34592E+01  AVG DEVIATION= 0 73418E+04

 THE CHI SQUARE STATISTIC OBTAINED DUE TO
 THE PRESENCE OF OUTLIERS USING MM IS '

CHI-SQUARE STATISTIC
********************

  4 538124

NO  OF DEGREES OF FREEDOM
*************************

          3

 THE THEORITICAL CHI-SQUARE VALUE IS '    7.815000

 FOR             3   DEGREES OF FREEDOM

 THE FIT IS GOOD

 THE CHI SQUARE STATISTIC OBTAINED WITHOUT
 THE PRESENCE OF OUTLIERS USING MM IS

CHI-SQUARE STATISTIC
********************

  4 777685

NO  OF DEGREES OF FREEDOM
*************************

          3

 THE THEORITICAL CHI-SQUARE VALUE IS '    7.815000

 FOR             3   DEGREES OF FREEDOM

 THE FIT IS GOOD
```

```
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA  ** PLOTTING RESULTS                           A
A----   - ---- -----------------------------------A
A    1 FREQUENCY PLOT                             A
A    2 C D F PLOT                                 A
A    3 CONFIDENCE BANDS                           A
A    4 TO MASTER MENU                             A
A       ---- --- ----  ------  - ---              A
Type the serial number of desired option
3

         NORMAL AND LOG NORMAL DISTRIBUTIONS
         AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

         PEARSON AND LOG PEARSON DISTRIBUTION
        AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 MEAN=      4 3693
 VAR=   0 32390E-01
 SKEW=   0 10856E 01
 STDEV=   0 17997E+00
 KURTOSIS=   0 28241E+01  AVG DEVIATION= 0 14389E+00
```

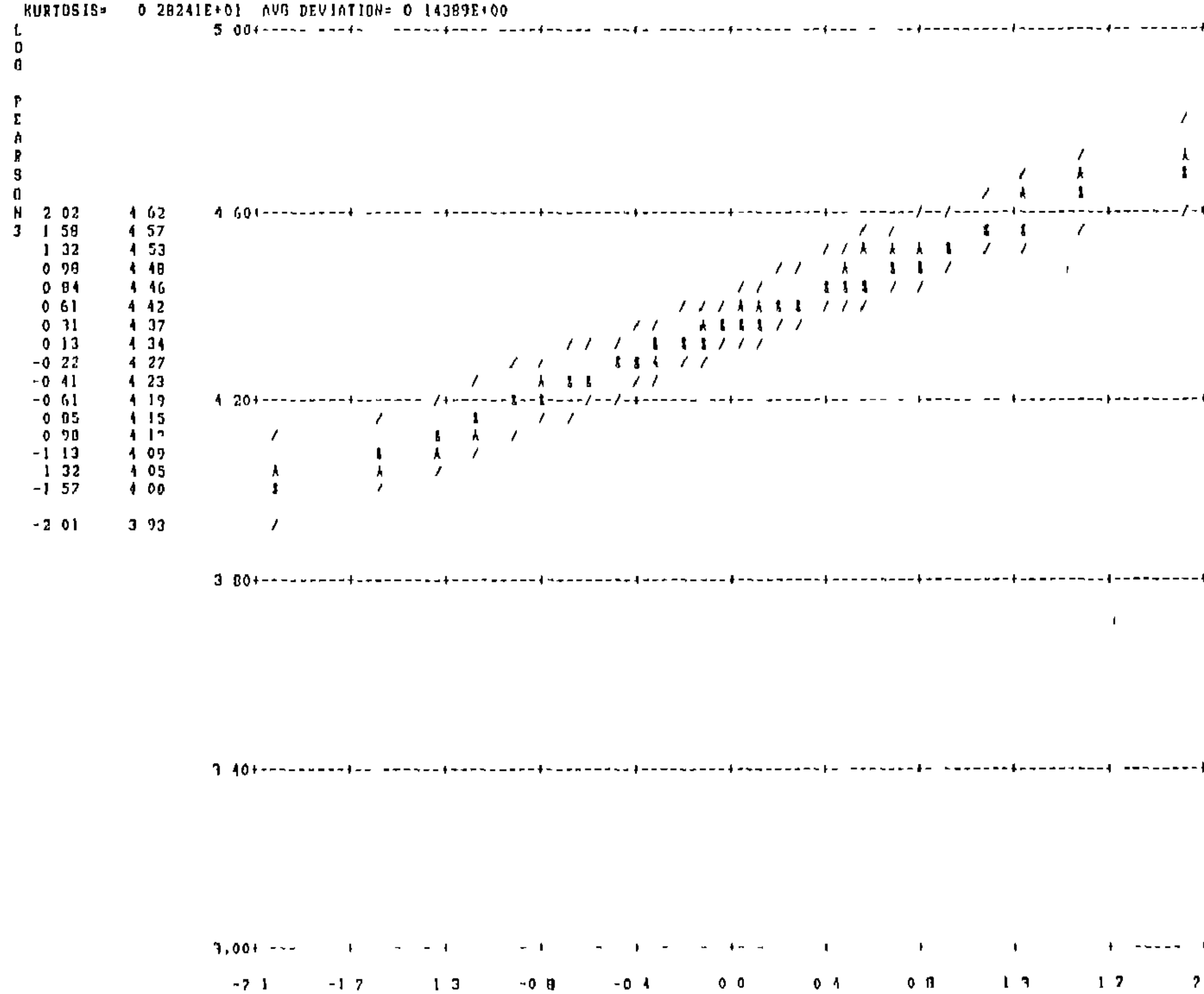


Fig 1

```
*-----------------------------------------------*
*      MASTER-MENU                              *
*-----------------------------------------------*
*   1 PRELIMINARY ANALYSIS                      *
*   2.CHOICE OF DISTRIBUTIONS                   *
*   3 TESTS FOR GOODNESS OF FIT                 *
*   4 PLOTTING OF RESULTS                       *
*   5.QUIT                                      *
*   6 HELP FROM EXPERT SYSTEM INTERFACE         *
*-----------------------------------------------*
VALUE PASSED IS    15 00000
NAME PASSED IS
PT3,LF3,LN3
Type the serial number of desired option
5
FORTRAN STOP
$
```

preliminary analysis is done and hence 1 is entered. The details of the Submenu of preliminary analysis are displyed and Statistical Parameters option is now chosen The statistical parameters, viz , the mean, variance, skew, standard deviation, kurtosis and average deviation for the original data as well as log transformed data are calculated The statistical analysis show that the data are positively skewed (Ø 9599), coefficient of kurtosis is 4 Ø343 and the coefficient of skew of the log transformed data is nearly zero (Ø ØØ96), the mean and standard deviation of the original data and log transformed data are 25382 2852, Ø 1Ø7Ø6EØ5 and 1Ø Ø593; Ø 41498 respectively Now, the preliminary advice is taken from the Expert system interface by entering 6 and then 1 The expert system asks for the type of data and whether outliers are present Depending on whether the data are exceedence, annual or seasonal, the letter e, a or s is typed Here the data being annual, a is entered Initially, since it is not known whether the outliers are present or not, nk is entered The expert system suggests that in the absence of information, outliers are initially assumed not to be present It suggests method of moments for parameter estimation and gives the list of steps for analysis, viz , preliminary analysis, test for outliers, fit a suitable distribution, test for goodness of fit and plot the results.

The KBES suggests all the possible distributions that can be fitted to the data For the dta of Narmada river, the KBES suggests a distribution which can be a Pearson Type 3, 3 parameter Log Normal distribution or Log Pearson Type 3 distribution

The analysis is done in the same sequence Initially the outlier tests are done for Pearson Type 3 distribution which indicates the presence of an upper outlier The Expert system interface is again approached and the secondary advice is taken Here, when the question ' Do you think outliers are present ' is asked, yes is entered. The expert system suggests method of least squares for parameter estimation Now, choice of distribution option is chosen by entering 2 and the method of least squares is chosen as the method of analysis by entering 3 The distribution fit is the Pearson Type 3 which is chosen by entering 4. In the method of least squares, the value of the oultier is neglected

while the plotting position value is retained and then the regression is done between the variable and the frequency factor Here, the rank (I), variable X(I), logarithm of the variable LX(I), the plotting position P(I), frequency factor for the variable and logarithm of the variable KTP(I) and KTLP(I) respectively) are displayed on the screen The Chi square statistic obtained with the deletion of outlier values using MLS is 1 3074 as against the theoretical value of 7 815 for 3 degrees of freedom and the fit is good The parameters for the Pearson type 3 distribution are displayed on the screen as follows

ALPHA = 4456 367	MEAN = 24320 04
BETA = 4 3412	STDEV = 9285 079
GAMMA = 4974 078	SKEW = 0 9599

The coefficient of skew of the original data is used to calculate the parameters of the Pearson Type 3 distribution The optimal value of the coefficient of skew can be obtained by conducting a linear search by Golden Search method, get a value of skew, then calculate K_T followed by a regression between X and K_T and then calaculate the sum of the squared errors, which is repeated until the sum of the squared errors is a minimum This has not been implemented in the study

Next the Chi square test is chosen This option calculates the Chi square value obtained with the deletion of outlier values from the distribution using method of moments parameter estimates, (1 3074) and the Chi square value obtained for the original data set using method of moments parameter estimates (2 1097) as against a theoretical value of 7 815 for 3 degrees of freedom

The above procedure is repeated for 3 parameter Log Normal distribution The outlier tests indicate the absence of any outlier and even though MLS option is available, the method of moments is adopted as the method of analysis The Chi square statistic obtained is 5 666 as against a theoretical value of 7 815 for 3 degrees of freedom The fit is hence good at 95 % confidence limit

The analysis for Log Pearson Type 3 indicates the presence of an upper outlier. So, MLS is used for the method of analysis The Chi square value obtained by MLS deleting the

outlier value is Ø 3843 as against a theoretical value of 7 815 for 3 degrees of freedom The parameters obtained for Log Pearson Type 3 are displayed on the screen as follows

ALPHA	= 8 111E-Ø4	MEAN	= 4 3552
BETA	= 43229 Ø7	STDEV	= Ø 1686
GAMMA	= -3Ø 7Ø81	SKEW	= 9 6193E-Ø3

Similarly the Chi square values obtained using method of moments with the deletion of outlier value from the data set and for the original data set are 4 5381 and 4.7777 respectively as against the theoretical value of 7 815 for 3 degrees of freedom at 95% confidence limit Since the Chi square value obtained by MLS which takes the outliers into consideration is the least, these parameters are taken to draw the confidence bands Now, on entering the number 3 for confidence bands, the fitted distribution indicated by '&' , the sample data indicated by '*' , and the confidence limits indicated by '/' are drawn (Fig 1 , Table 5 2)

Since the Chi square value obtained by the deletion of the outlier value from the data set using the method of least squares for parameter estimation is the least for Log Pearson Type 3 distribution, this distribution is taken as the best fit for the data of river Narmada among the distributions tested for the methods of parameter estimates used The Histogram for the data is given in Fig 5 1A and the cdf plot is given in Fig 5 1B The summary of the analysis is given in Table 5 3

ii) Set 2 The analysis for the monsoon flows for Hirakud shows that the data are positively skewed (1 1925) and a coefficient of kurtosis of 6 8911. The KBES suggests a distribution which is either a Pearson Type 3 or Log Pearson Type 3 or 3 parameter Log Normal distribution

The Pearson type 3 distribution has an upper outlier and no lower outliers The chi square statistic obtained by taking the outlier into consideration and using method of least squares as the method of analysis is 6 8949 The Chi square values obtained by using the method of moments with the outlier and without outlier are 6 89 and 11 49 respectively as against a theoretical value of 7 815 for 3 degrees of freedom Except for the Chi square value obtained for the original data using MM which is high, the

```
           |
           |
           |
           |
           |
           |
           |
           |
F    9-+
R    8-+
E    7-+    +-----+
Q    6-+-----+    +-----+
U    5-+     |    |     +-----+
E    4-+     |    |     |     |
N    3-+     |    |     |     |
C    2-+     |    |     |     |
Y    1-+     |    |     |     +-----+-----+-----+-----+
       +-----+----+-----+-----+-----+-----+-----+-----+-------------
       11         23          35          47          59
```

FIG. 5.1A HISTOGRAM FOR FLOOD PEAKS IN RIVER NARMADA

Fig 5 1B CDF Plot for Flood Peaks in River Narmada

LOG PEAKS ON 3	
4 73	0 98
4 68	0 94
4 62	0.91
4 56	0 87
4.53	0.84
4 52	0 80
4 50	0 77
4 50	0 73
4 47	0 69
4 45	0 66
4 42	0 62
4.41	0 59
4 40	0 55
4.39	0 52
4 37	0 48
4 36	0 45
4 32	0 41
4 29	0 38
4 29	0 34
4 27	0 31
4 26	0 27
4 25	0 23
4 22	0 20
4 21	0 16
4 13	0 13
4 08	0 09
4 05	0 06
4 05	0 02

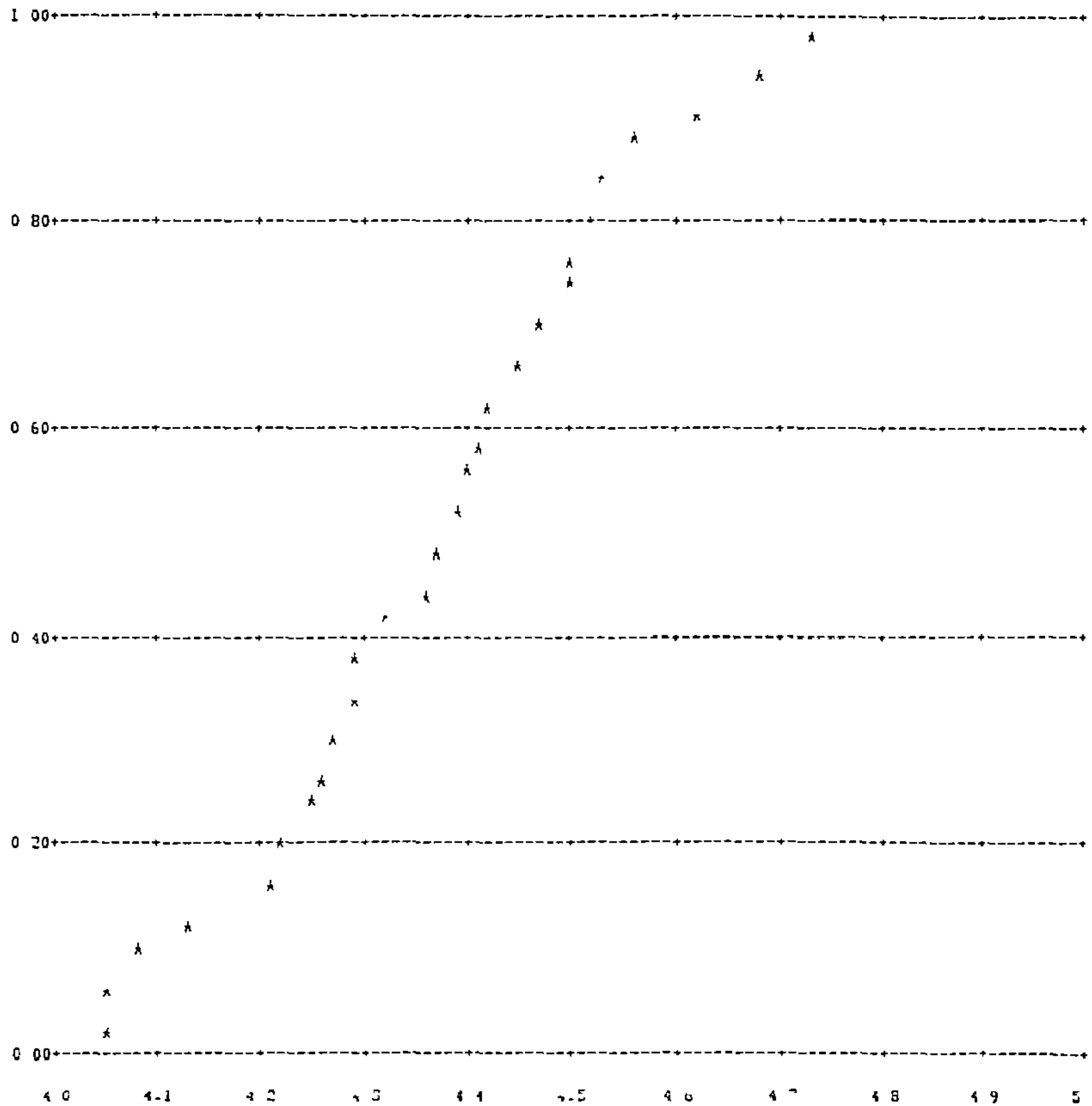


Table 5 3

Analysis for river Narmada

Theoretical Chi square value χ^2(95% , 3DOF) = 7 815

Stat parameters		Dist suggested	Outliers if any		Outliers considered		Method of analys	Dist fit	χ^2 valu
kurt	skew		high	low	High	Low			
4 Ø3	Ø 959	PT3	1	Ø	1	Ø	MLS		1 31
					1	Ø	MM		1.31
					Ø	Ø	MM		2 11
		LN3	Ø	Ø	Ø	Ø	MM		5 66
		LPT3	1	Ø	1	Ø	MLS	LPT3	Ø 38
					1	Ø	MM		4 54
					Ø	Ø	MM		4 77

Distribution Fit — Log Pearson Type 3

Method of analysis — Method of least squares

Estimated Parameters — ALPHA = 8 11E-Ø4

BETA = 43229 Ø7

GAMMA = -3Ø 7Ø81

other two Chi square values obtained using MM (deleting the outlier value) and MLS (deleting the outlier value) indicate the Pearson Type 3 distribution is a good fit at 95% confidence level

The analysis for 3 parameter Log Normal distribution does not indicate any outliers The parameters are estimated using the equations given by Kite (1977) A subroutine to estimate the parameters is also available which has been incorporated in the study The Chi square value is 5 83. The analysis for Log Pearson Type 3 indicates the presence of an upper outlier The Chi square value obtained using the method of least squares taking the outliers into consideration is 8 945 The Chi square values obtained using the method of moments with the outliers taken into consideration and due to the absence of outliers are 6 89 and 11 49 respectively

The 3 parameter Log Normal distribution is the best fit with a Chi square value of 5 83 χ^2(95%, 3DOF) = 7 815 The parameters for the 3 parameter Log Normal distribution obtained is given in Table 5 4 The T year recurrence interval estimates are also given in the Fig 5 4 The histogram for the data and the cdf plot are given in Fig 5 2A and Fig 5 2B The results are shown in Fig 5 3.

5.2.2 Seasonal data - The analysis for all the seasons for the streamflow data for Mahanadi basin is done and the analysis for the 3rd season, i e , for the month of August is reported completely in Table 5 5 The histogram and c d f plot are shown in the Figs 5 4A and 5 4B The fitted distribution is shown in Table 5 6 The T year recurrence events are also shown in the Table 5 6

The analysis for the month of August (3rd season) shows that the data are negatively skewed (-0 0934) and the coefficient of kurtosis is 2 6669 The KBES suggests Normal distribution for the data Outlier analysis indicates the absence of any outliers for the Normal distribution The KBES suggests the method of moments as the method of analysis Using the method of moments, the Normal distribution is fitted The observed Chi square statistic is 1 163. χ^2(95%, 3DOF) = 7 815. The confidence bands plot shows that the Normal distribution is the best fit

The analysis for each season, i.e , the distribution suggested and the distribution fit are given in Table 5 7

TABLE 5 4

FITTED DISTRIBUTION FOR MONSOON FLOWS FOR RIVER MAHANADI AT HIRAKUD

THREE PARAMETER LOGNORMAL DISTRIBUTION

METHOD OF MOMENTS

MEAN OF X	0.36383E+04
VARIANCE OF X	0.21487E+07
SKEW OF X	0.11436E+01
A	-0 37778E+03

T,YEARS	2	5	10	20	50	100
X_T	0.33948E+04	0.47026E+04	0.55580E+04	0 63718E+04	0 74219E+04	0.82113E+04
S_T	0.27421E+03	0.35519E+03	0.47614E+03	0 70147E+03	0.11334E+04	0.15421E+04

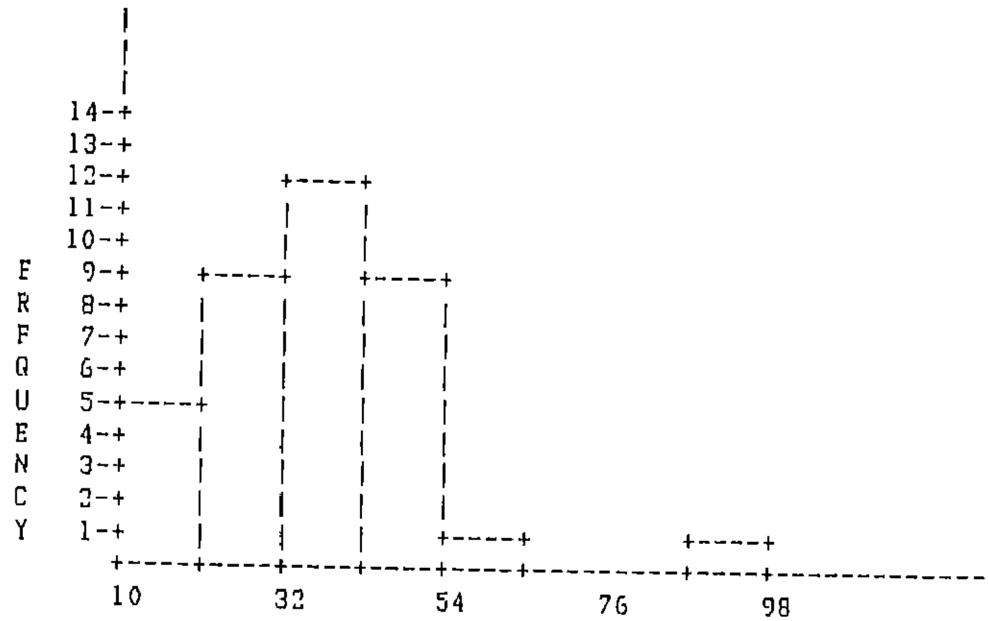


FIG. 5 2A HISTOGRAM FOR MONSOON FLOWS IN HIRAKUD

Fig 5 2B CDF Plot for Monsoon Flows in Hirakud

LOG NORMAL EX	
9 11	0 98
8 68	0 96
8 57	0 93
8 53	0 90
8 46	0 88
8 44	0.85
8 42	0 82
8 41	0 80
8 41	0 77
8 40	0 74
8 40	0 71
8 34	0 69
8 33	0 66
8 32	0 63
8 30	0.61
8 30	0 58
8 29	0 55
8.27	0 53
8 23	0 50
8 22	0 47
8 21	0 45
8 16	0 42
8 11	0 39
8 05	0 37
7 99	0.34
7 84	0 31
7 83	0.29
7.82	0 26
7 76	0 23
7 75	0 20
7 74	0 18
7 73	0 15
7 59	0 12
7 55	0 10
7 44	0.07
7 31	0 04
7 15	0 02

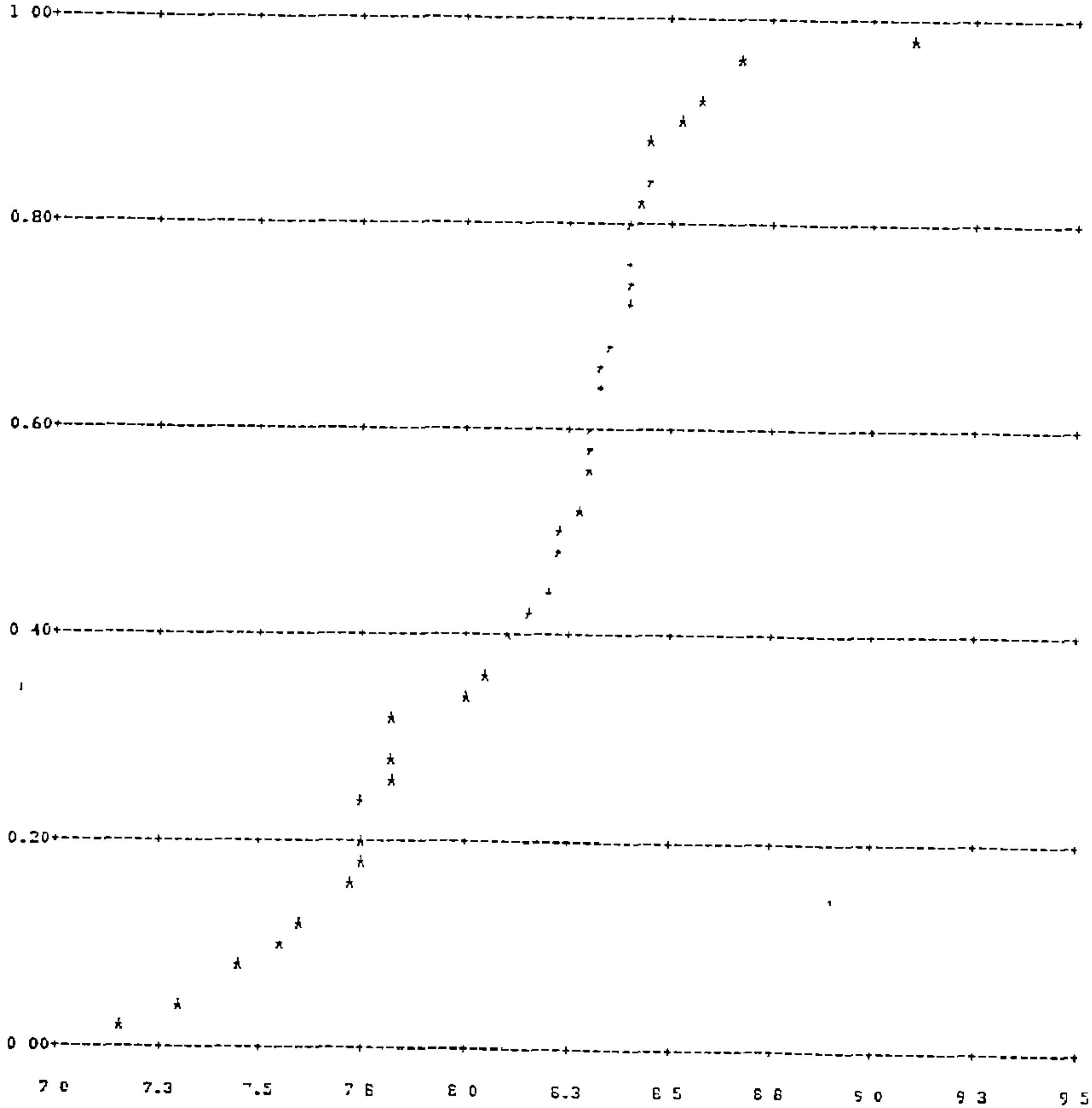


Fig 5 3 Three Parameter Log Normal Distribution for Nonseasonal Data in Hirakud

LOG NORMAL

1 94	8 80
1 62	8 69
1 41	8.63
1 25	8 57
1.00	8.49
0 90	8.46
0 80	8 42
0 63	8 36
0 48	8 31
0 34	8 26
0 20	8 21
0.07	8 16
-0 13	8 08
-0 27	8 03
-0 41	7 98
-0 48	7 95
-0.63	7 88
-0 72	7 85
-0 80	7 81
-1 00	7 73
-1 12	7 68
-1 25	7 62
-1 41	7 56
-1 62	7 47
-1 94	7 33

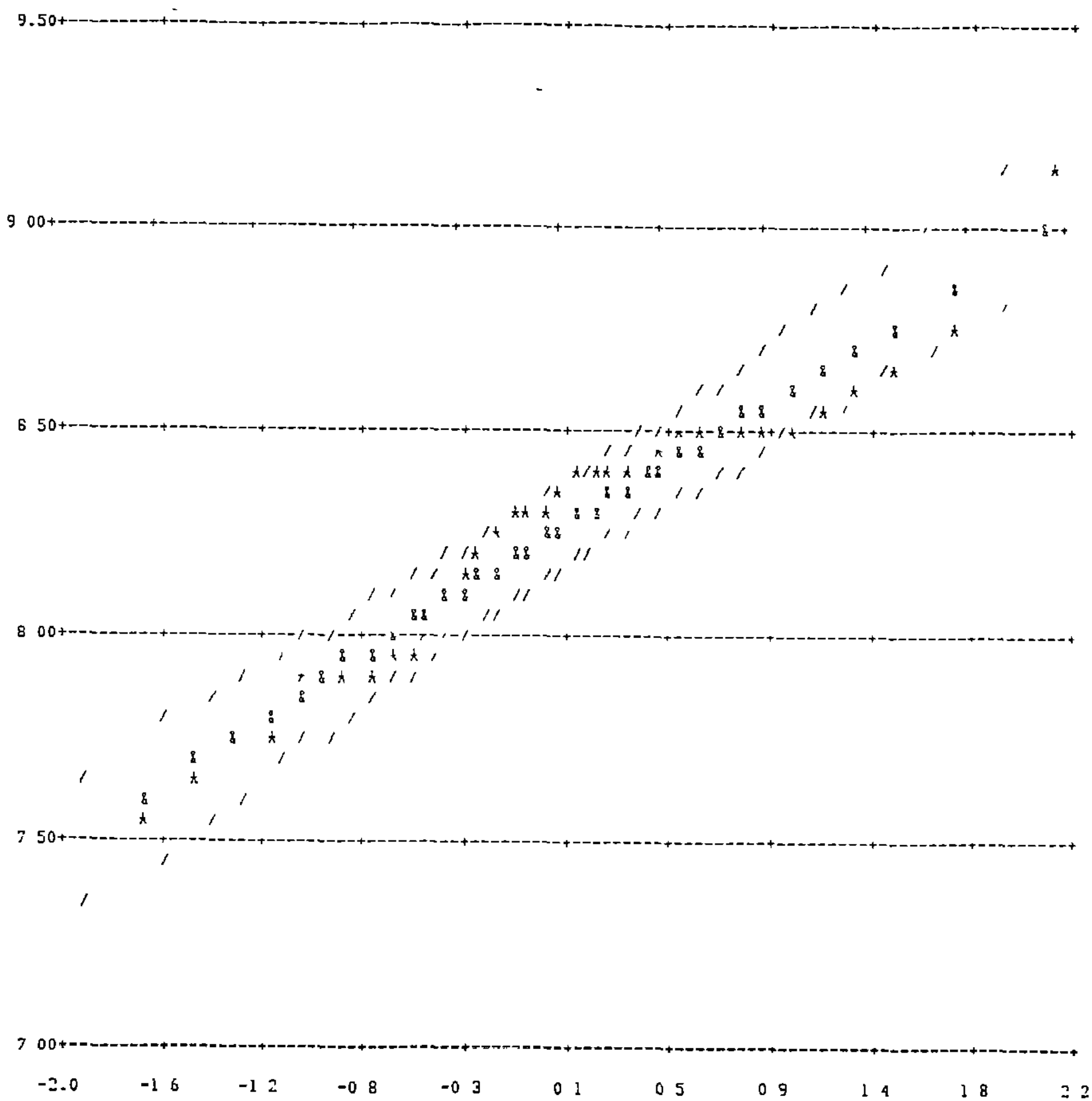


Table 5 5 Interactive Session II
(Seasonal Data for the Month of August for Mahanadi Basin)

```
*********************************************
*      EXPERT SYSTEM FOR FREQUENCY ANALYSIS    **
*            OF HYDROLOGIC DATA               **
*               BY O P MISHRA                 **
*            IMPROVED BY K.R C RAO            **
*          UNDER  THE  GUIDANCE  OF           **
*             Dr S Ramaseshan                 **
*             at I.I T.Kanpur                 **
*********************************************
*    WISH   YOU   HAPPY COMPUTING             **
*              W I T H                        **
*         F A C H V E S Ver 2.0               **
*********************************************
* *--* FREQUENCY-ANALYSIS WITH EXPERT ADVICE *--*
* ---------------------------------------------*
**************************************
*    I N P U T   M O D U L E         *
* -----------------------------------*
Do you want to give input on the screen?
If yes, Type 1 ,If no, Type 2
   If you want to quit Type 3
2
--The input file should be named as TEST.INP
--        It must have following :
--       *   a TITLE with format 80A1
--             AND, unformated
--       *   NX--Number of data
--       *   NS--Number of seasons
--         (for annual series NS =1)
--       *   NPT--Number of points
--           NPT=NCLASS+1 where NCLASS is number of
               classes in which data can be divided
               for chisquare test  The class frequency
               should be equal to 5
--            (It can be 7/11/13)
--       *   Observational hydrologic data RX(I)
--       (Not more than 600;
                seasons(in row)/years(in column)
Have you got such an input file ? If yes,type 1
       If you want to quit ,type 2
       If you want to go to previous menu type 3
1
For which season do you want to do the analysis ?
Type the serial no of the season
3
   THIS COMPUTATION IS FOR SEASON NO          3
*---------------------------------------------*
*      MASTER-MENU                             *
*---------------------------------------------*
*   1.PRELIMINARY ANALYSIS                     *
*   2.CHOICE OF DISTRIBUTIONS                  *
*   3 TESTS FOR GOODNESS OF FIT                *
*   4.PLOTTING OF RESULTS                      *
*   5.QUIT                                     *
*   6 HELP FROM EXPERT SYSTEM INTERFACE        *
*   7.ANALYSIS FOR OTHER SEASONS               *
*---------------------------------------------*
Type the serial number of desired option
```

```
1
****************************************
*   $$$ PRELIMINARY ANALYSIS $$$       *
*--------------------------------------*
*   1.STATISTICAL PARAMETERS           *
*   2.FREQUENCY PLOT                   *
*   3 C D F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                 *
*   5.TRANSFORMATIONS                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE *
*   7 TO MASTER MENU                   *
*   8 OUTLIER TESTS                    *
*--------------------------------------*
Type the serial number of desired option
1
 MEAN=  15156.8398
 VAR=   0 18210E+08
 SKEW=  -0 93395E-01
 STDEV=   0.42674E+04
 KURTOSIS=   0.26669E+01  AVG DEVIATION= 0 34710E+04

THIS ANALYSIS IS FOR LOG-TRANSFORMED DATA
 MEAN=       9 5821
 VAR=   0 97444F-01
 SKEW=  -0 78038E+00
 STDEV=   0.31216E+00
 KURTOSIS=   0.33245E+01
XMIN= 0 87581E+01  XMAX= 0 10099E+02

****************************************
*   $$$ PRELIMINARY ANALYSIS $$$       *
*--------------------------------------*
*   1.STATISTICAL PARAMETERS           *
*   2.FREQUENCY PLOT                   *
*   3.C.D.F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                 *
*   5.TRANSFORMATIONS                  *
*   6 HELP FROM EXPERT SYSTEM INTERFACE *
*   7 TO MASTER MENU                   *
*   8.OUTLIER TESTS                    *
*--------------------------------------*
Type the serial number of desired option
6
****************************************
*  $$ EXPERT SYSTFM INTERFACE $$      **
*--------------------------------------*
*   1 FOR PRELIMINARY ADVICE           *
*   2 FOR SECONDARY  ADVICE            *
*   3 FOR EXPERT ANALYSIS OF RESULTS   *
*   7 TO MASTER MENU                   *
*--------------------------------------*
Which help do you want?
Type the serial number of desired option
```

```
1
f-0     (initial-fact)
f-1     (start)
f-0     (initial-fact)
f-1     (start)
f-2     (mean 15157)
f-3     (skew -0.093395)
f-4     (kurt 2 66689992)
f-5     (stdev 4267 39990234)
f-6     (stdev of logtran data 0 31215999)
f-7     (length of data 50)
       W E L C O M E

           TO

        E A C H V E S

This expert   system   is  designed  for  single
             peaked  continuous hydrologic variables   only

IF YOU DO NOT UNDERSTANT ANY QUESTION ; TYPE nk
            nk --- Not known

Type of data ---- type the first letter (e/a/s)
             Whether it is exceedence OR annual ?a
Do you think that outliers are present?  nk
NORMAL DISTRIBUTION CAN BE FIT
In absence of information outliers are
             initially assumed not to be present

Use method of moments for parameter estimation

Here is the list of steps for analysis     **
             1 **DO THE PRELIMINARY ANALYSIS
             2 **TEST FOR OUTLIERS
             3 **FIT A SUITABLE DISTRIBUTION
             4.**TEST FOR GOODNESS OF FIT
             5 **PLOT THE RESULTS
1.You can fit either one OR all available distributions

 ** If more than one distribution satisfies the
               goodness of fit test (chi square test) then
               the best fit is the one with the least
               Chi square value

PROCESS IS BEING STOPPED;
             RESET THE PROCESS IF YOU WANT TO CONTINUE
             WITH OTHER SET OF DATA
Do you want any help ?
```

```
no
Suggestions passed should be taken in sequencial order
f-0     (initial-fact)
f-1     (start)
f-2     (mean 15157)
f-3     (skew -0 093395)
f-4     (kurt 2.66689992)
f-5     (stdev 4267 39990234)
f-6     (stdev of logtran data 0.31215999)
f-7     (length of data 50)
f-8     (type of data a)
f-9     (process continue yes)
f-10    (data is large enough)
f-11    (outliers are ok)
f-12    (suggestion 6)
f-13    (esad ND)
f-14    (outliers no)
f-15    (stop)
f-16    (help no)
Excising rule' rule1
Excising rule' rule2
Excising rule' rule3
Excising rule' rule4
Excising rule' rule5
Excising rule: rule6
Excising rule' rule7
Excising rule  rule8
Excising rule  rule9
Excising rule' rule10
Excising rule' rule11
Excising rule  rule12
Excising rule: rule13
Excising rule  rule14
Excising rule  rule15
Excising rule' rule16
Excising rule  rule18
Excising rule. rule19
Excising rule  rule20
Excising rule. rule21
Excising rule  rule22
Excising rule' rule23
Excising rule  rule24
Excising rule' rule25
Excising rule  rule26
Now you are out of clips
Value passed =    6.000000
Name passed is =
ND
*-----------------------------------------------*
*       MASTER-MENU                             *
*-----------------------------------------------*
*   1 PRELIMINARY ANALYSIS                      *
*   2.CHOICE OF DISTRIBUTIONS                   *
*   3 TESTS FOR GOODNESS OF FIT                 *
*   4.PLOTTING OF RESULTS                       *
*   5.QUIT                                      *
*   6 HELP FROM EXPERT SYSTEM INTERFACE         *
*   7 ANALYSIS FOR OTHER SEASONS                *
*-----------------------------------------------*
VALUE PASSED IS    6.000000
NAME PASSED IS
ND
Type the serial number of desired option
```

```
1
****************************************
*   $$$ PRELIMINARY ANALYSIS $$$        *
*---------------------------------------*
*   1 STATISTICAL PARAMETERS            *
*   2.FREQUENCY PLOT                    *
*   3.C.D.F PLOT (FOP ALL DISTRIBUTIONS) *
*   4.CONFIDENCE BANDS                  *
*   5.TRANSFORMATIONS                   *
*   6 HELP FROM EXPERT SYSTEM INTERFACE *
*   7 TO MASTER MENU                    *
*   8.OUTLIER TESTS                     *
*---------------------------------------*
Type the serial number of desired option
8
 MEAN=  15156 8398
 VAR=   0.18210E+08
 SKEW=  -0.93395E-01
 STDEV=   0.42674E+04
 KURTOSIS=   0.26669E+01  AVG.DEVIATION= 0.34710E+04

 THE DATA IS BEING CHECKED FOR OUTLIERS
 FOR WHICH DISTRIBUTION DO YOU WANT IT CHECKED i.e
 WHETHER 1  NORMAL,  2. LOG-NORMAL(2 PARAMETERS)
 3  LOG-NORMAL(3 PARAMETERS),  4  EV-I
 5. EV-III,  6. PT-III,  7. LPT-III
1

 THE PARAMETEPS FOR THE ORIGINAL DATA ARE ·
 MEAN=  15156 8398
 SKEW=  -0 93395E-01
 STDEV=   0 42674E+04
 *** FOR ORIGINAL DATA ***
CHECK FOR OUTLIERS BY WRC METHOD

 THE PARAMETERS FOR THE LOG-TRANSFORMED DATA ARE ·
 MEAN=       0 0000
 SKEW=   0 00000E+00
 STDEV=   0.00000E+00

 XMIN=  0.63630E+04     XMAX=  0 24313E+05
 FOR NORMAL DISTRIBUTION
 THE THRESHOLD VALUE FOR HIGHER OUTLIERS      26968 90
 THE THRESHOLD VALUE FOR LOWER OUTLIERS ·    3344 778
XMIN= 0 63630E+04  XMAX= 0.21313E+05
NO OF HIGH OUTLIERS =      0
NO.OF LOW OUTLIERS=     0
 OUTLIERS ARE ABSENT FOR THIS DISTRIBUTION
****************************************
*   $$$ PRELIMINARY ANALYSIS $$$        *
*---------------------------------------*
*   1 STATISTICAL PARAMETERS            *
*   2 FREQUENCY PLOT                    *
*   3 C D.F PLOT (FOR ALL DISTRIBUTIONS) *
*   4 CONFIDENCE BANDS                  *
*   5 TRANSFORMATIONS                   *
*   6.HELP FROM EXPERT SYSTEM INTERFACE *
*   7.TO MASTER MENU                    *
*   8.OUTLIER TESTS                     *
*---------------------------------------*
Type the serial number of desired option
```

```
7
*---------------------------------------------------*
*       MASTER-MENU                                 *
*---------------------------------------------------*
*    1 PRELIMINARY ANALYSIS                         *
*    2 CHOICE OF DISTRIBUTIONS                      *
*    3.TESTS FOR GOODNESS OF FIT                    *
*    4 PLOTTING OF RESULTS                          *
*    5 QUIT                                         *
*    6 HELP FROM EXPERT SYSTEM INTERFACE            *
*    7 ANALYSIS FOR OTHER SEASONS                   *
*---------------------------------------------------*
VALUE PASSED IS     6 000000
NAME PASSED IS
ND
Type the serial number of desired option
2
************************************************
*  $$ METHOD OF ANALYSIS $$           *
*-------------------------------------*
*    1 METHOD OF MOMENTS              *
*    2 METHOD OF MAXIMUM LIKELYHOOD   *
*    3.METHOD OF LEASTS-SQUARES       *
*    4 TO MASTER MENU                 *
*-------------------------------------*
Which method do you like?
Type the serial number of desired option
1
******************************************
**    $$ CHOICE OF DISTRIBUTION $$      **
*---------------------------------------*
*    1 NORMAL                    *
*    2 LOG-NORMAL (2-PARAMETER)  *
*    3 LOG-NORMAL (3-PARAMETER)  *
*    4 PEARSON TYPE 3            *
*    5 LOG-PEARSON TYPE-3        *
*    6 EXTREME-VALUE TYPE-1      *
*    7 EXTREME-VALUE TYPE-3      *
*    8 FAP--PROGRAMS             *
*    9 TO MASTER MENU            *
*   10 TRANSFORMATIONS           *
*   11 METHOD OF ANALYSIS        *
*-----------------------------------------*
If you want to transform the data**TYPE-10
If not,then which distribution? type the serial number
1
**************************************************
*  $$ TEST OF GOODNESS OF FIT $$             **
*---------------------------------------------*
*     1 CHI-SQUARE TEST                       *
*     2.KOLMOGOROV-SMIRNOV TEST               *
*     3 TO PLOTTING RESULTS                   *
*     4 COMPARISION OF STANDARD ERROR         *
*       OF ESTIMATE FOR DIFFERENT             *
*         DISTRIBUTIONS                       *
*     5 TO MASTER MENU                        *
*     6 HELP FROM EXPERT-SYSTEM INTERFACE     *
*---------------------------------------------*
Type the serial number of desired option
```

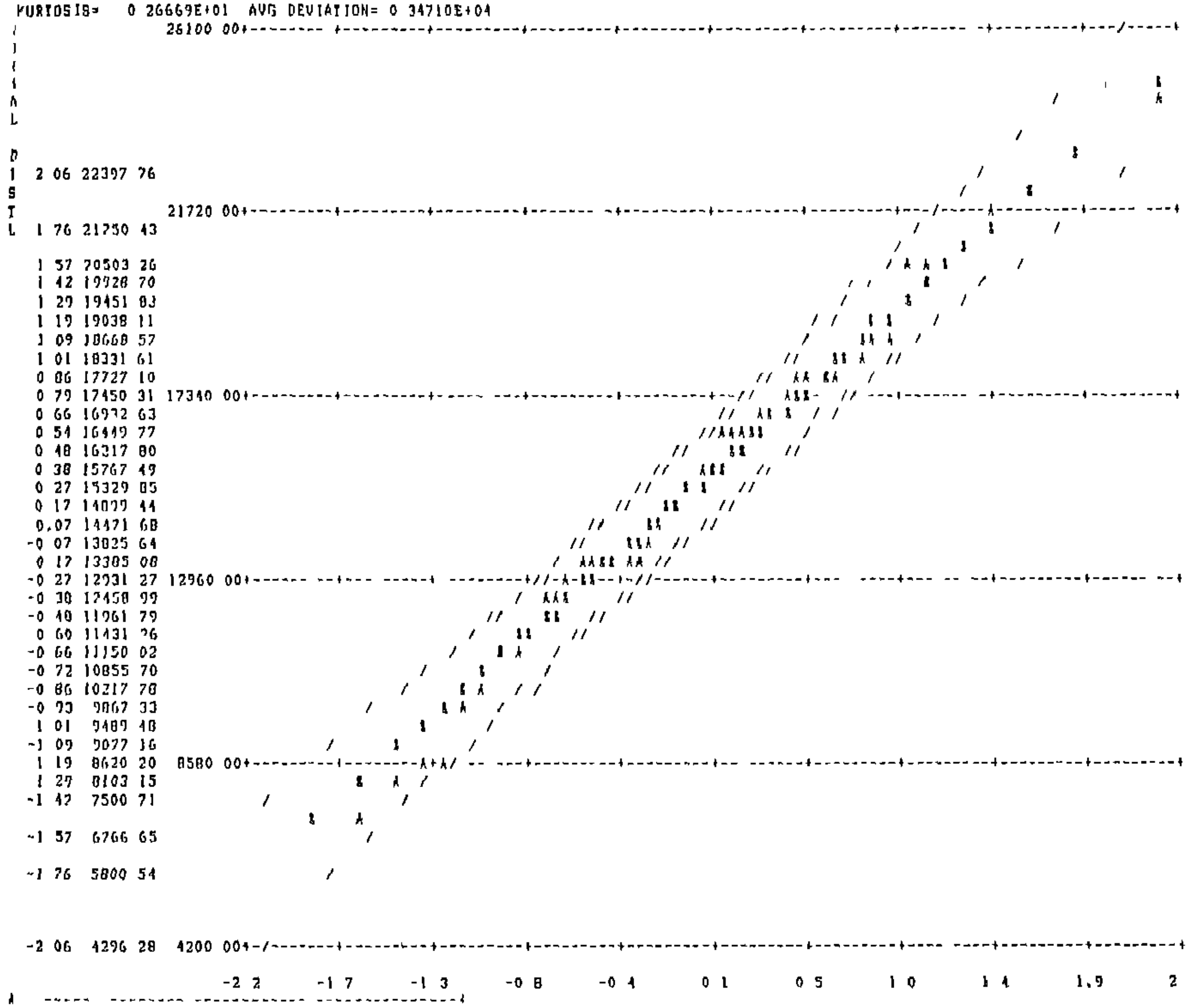
```
I

CHI-SQUARE STATISTIC
********************

  1 162785

NO  OF DEGREES OF FREEDOM
*************************

          3

 THE THEORITICAL CHI-SQUARE VALUE IS :    7 815000

 FOR            3   DEGREES OF FREEDOM

 THE FIT IS GOOD
**********************************************
**  $$ PLOTTING RESULTS                *
*--------------------------------------*
*   1.FREQUENCY PLOT                   *
*   2.C.D.F.PLOT                       *
*   3.CONFIDENCE BANDS                 *
*   4.TO MASTER MENU                   *
*--------------------------------------*
Type the serial number of desired option
```

NORMAL AND LOG NORMAL DISTRIBUTIONS
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

NORMAL AND LOG NORMAL DISTRIBUTIONS
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

EAN= 15156 8398
AR- 0 18210E+08
KEW -0 93395F 01
STDEV= 0 42674F+04
URTOSIS- 0 26669F+01 AVG DEVIATION= 0 34710E+04

MEAN= 15156 8398
VAR- 0 18210E+08
SKEW= -0 93395F 01
STDEV= 0 42674E+04
KURTOSIS= 0 26669E+01 AVG DEVIATION= 0 34710E+04



```
*      MASTER-MENU                                  *
*---------------------------------------------------*
*   1.PRELIMINARY ANALYSIS                          *
*   2.CHOICE OF DISTRIBUTIONS                       *
*   3.TESTS FOR GOODNESS OF FIT                     *
*   4.PLOTTING OF RESULTS                           *
*   5.QUIT                                          *
*   6.HELP FROM EXPERT SYSTEM INTERFACE             *
*   7.ANALYSIS FOR OTHER SEASONS                    *
*---------------------------------------------------*
VALUE PASSED IS    6.000000
NAME PASSED IS
ND
Type the serial number of desired option
5
FORTRAN STOP
$
```

```
       |
       |
    15-+
    14-+
    13-+                   +----+
    12-+                   |    |
    11-+         +----+    |    |
    10-+         |    |    |    |
 F   9-+         |    |    |    |
 R   8-+         |    +----+    |
 E   7-+         |    |    |    |
 Q   6-+         |    |    |    +----+
 U   5-+         |    |    |    |    |
 E   4-+----+----+    |    |    |    |
 N   3-+    |    |    |    |    |    +----+
 C   2-+    |    |    |    |    |    |    |
 Y   1-+    |    |    |    |    |    |    +----+
       +----+----+----+----+----+----+----+----+-------------
       6         11        16        21        26
```

FIG. 5.4A HISTOGRAM FOR SEASONAL DATA FOR THE MONTH OF AUGUST FOR MAHANADI BASIN

Fig 5 4B CDF Plot for Seasonal Data for the Month of August for Mananadi Basin

N		
O	24 31	0 99
R	22 85	0.97
M	21 94	0.95
A	21 57	0 93
L	20 67	0.91
	20 39	0 89
D	20 28	0 87
I	20 23	0 85
S	18 86	0 83
T	18 82	0 81
L	18 26	0 79
	18 07	0 77
	17 97	0 75
	17 66	0.73
	17 63	0 71
	17 59	0 69
	17 49	0 67
	17 07	0 65
	16 74	0 63
	16 58	0 61
	16 54	0 59
	16 36	0 57
	16 33	0 55
	15 74	0 53
	15 73	0 51
	14 76	0 47
	14 61	0 45
	14 45	0 43
	13 74	0 41
	13 59	0 39
	13 48	0 37
	13 36	0 35
	13 34	0 33
	13 32	0 31
	13 22	0 29
	12 97	0 27
	12 54	0 25
	12 38	0 23
	11 66	0 21
	11 35	0 19
	11 22	0 17
	10.24	0 15
	9 98	0 13
	8 77	0 11
	8 72	0 09
	8.27	0 07
	7 44	0 05
	7.35	0 03
	6 36	0 01

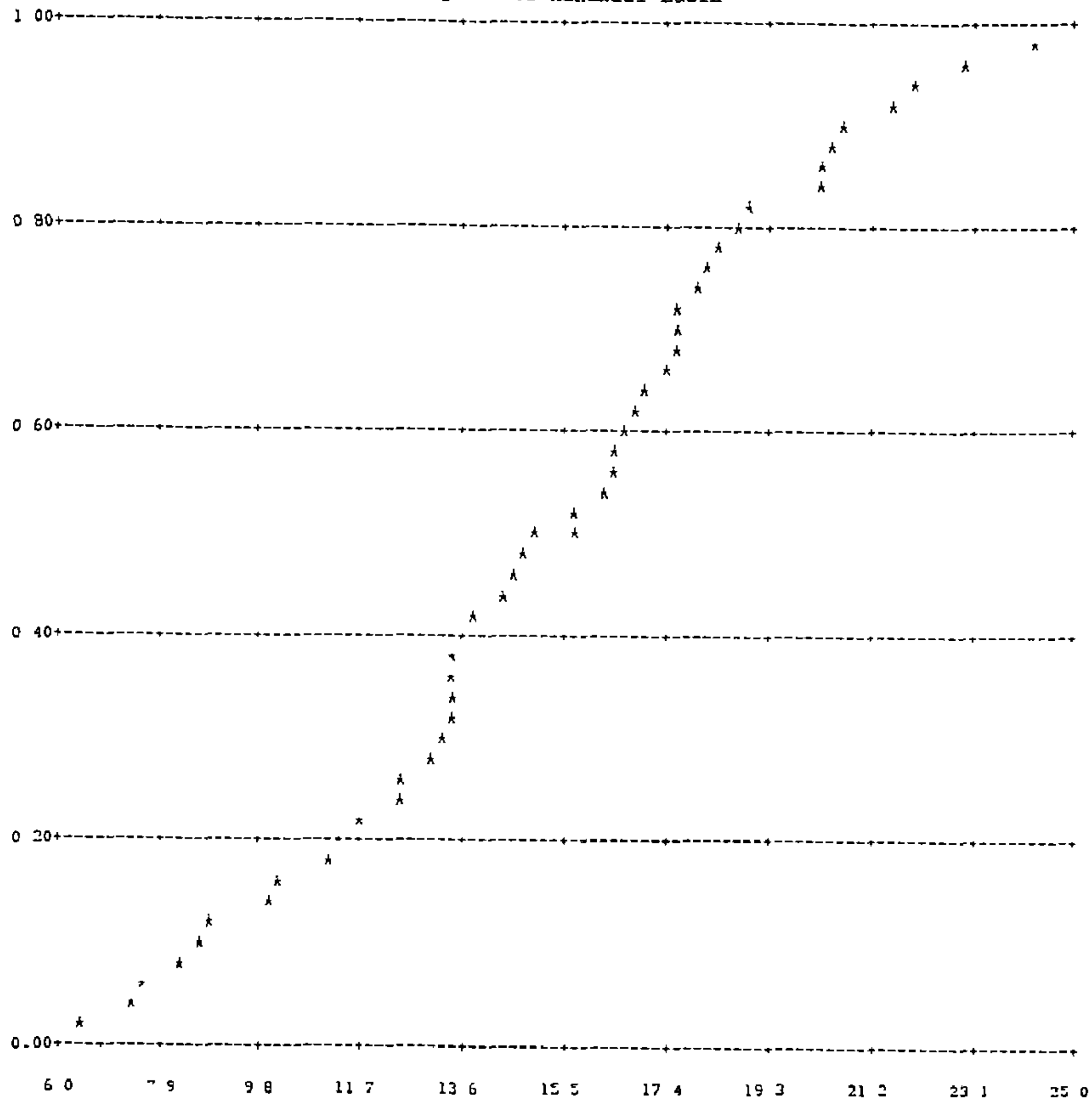


TABLE 5.6

PARAMETERS FOR NORMAL DISTRIBUTION FOR SEASONAL DATA IN MAHANADI BASIN

****
NORMAL DISTRIBUTION
****
MEAN= 15156 839844
STANDARD DEVIATION= 4267.363281
****

T YEARS	T YEAR EVENTS
2.0	15156.839844
5 0	18741.425781
10 0	20619 064453
20.0	22197.990234
50 0	23947 609375
100 0	25099.796875

Table 5 7

Analysis for Seasonal data for Mahanadi Basin

Theoretical Chi square value χ^2(95% , 3DOF) = 7 815

Season no	Stat parameters		Dist suggested	Outliers if any		Outliers considered		Method of analys	Dist fit Good	χ^2 valu
	kurt	skew		high	low	High	Low			
1	8 37	1 99	PT3	1	Ø	Ø	Ø	MM	PT3	3 36
			LN3	Ø	Ø					
			LPT3	1	Ø					
2	7 61	1 45	PT3	1	Ø	1	Ø	MLS	PT3	1 87
			LN3	Ø	Ø	Ø	Ø	MM	LN3	2 38
			LPT3	1	Ø	1	Ø	MLS	LPT3	1 37
3	2 66	-Ø Ø9	N	Ø	Ø	Ø	Ø	MM	N	1 16
			PT3	1	Ø	1	Ø	MLS	PT3	1 37
4	5 3Ø	Ø 892	PT3	1	Ø	1	Ø	MM	PT3	3 36
			LN3	Ø	3			&		
			LPT3	1	Ø			MLS		
5	5 24	1 14	LN2	Ø	Ø					
			PT3	2	Ø	Ø	Ø	MM	PT3	Ø.92
			LN3	Ø	Ø					
			LPT3	2	Ø					
6	6 31	Ø 628	PT3	2	Ø	2	Ø	MLS	PT3	2 57
			LN3	1	Ø					
			LPT3	Ø	1					

From considerations of consistency, PT3 is a good fit for all months and nonmonsoon season

5.3 Conclusions

The test runs of the program FACHVES Ver 2 Ø with the above set of data shows that the KBES is very helpful in fitting a probability distribution to a given set of data The distribution with the least chi square is taken as the best fit for the data set when only one nonseasonal data set is considered In case of seasonal data consistency of distributions over the seasons may also be considered They are not implemented in the program

CHAPTER 6

SUMMARY , CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

6.1 Summary

Frequency analysis of hydrologic data is a knowledge domain where lot of decisions have to be taken based on human expertise and intuition Frequency analysis being data dependent is subject to many errors Moreover, there is no fixed distribution that can be fitted to a given set of data An earlier study dealt with the structure and development of a ES based computer program FACHVES The Expert system is embedded in the FORTRAN program The program is written in FORTRAN language and the expert system has been developed for the VAX-VMS environment This study deals with modification and enhancement of capabilities of the earlier version

In the present version, a number of subroutines have been added which helps in a more detailed analysis of the data set For example, the subroutine OUTLIER identifies the number and the values of the outliers present in the data, the subroutine CONFBND is a plotting subroutine which plots the fitted distribution, the sample data and the confidence bands Other subroutines include FREQPL and TPLOT for plotting, MLS for method of least squares and GOFITST for comparing theoretical and actual Chi square values in the goodness of fit test In the given program i a distribution can be fitted to the data after taking Expert advice, ii the given set of data can be tested for outliers, iii if outliers are present, parameters can be re estimated by MLS, iv the goodness of fit tested by Chi square test, v the fitted distributions visualised graphically by seeing the plots and vi which distribution is more appropriate to the given set of data may then be decided Hence, the modifications done has enhanced the capabilities in fitting a distribution

6.2 Conclusions

Frequency analysis of hydrologic data is a knowledge domain where lot of decisions are to be taken based on human

intuition, experience and subjective judgement

The developed package of FACHVES Ver 2 Ø has been tested with three sets of data representing recorded streamflow data at two different sites in India The results of this study verify the satisfactory performance of the package , and in particular the enhancements of outlier test and graphic output

8.3 Suggestions for further study

A large amount of expert knowledge is required and available for frequency analysis of hydrologic data The package can be improved for a better practical use. Some of the modifications that can still be done are

1. Include more tests for testing outliers, e g , block procedures and robust estimates can be used
2. A linear Golden search method can be introduced to calculate the optimal value of coefficient of skew which is used to calculate the parameters of Pearson Type 3 or Log Pearson Type 3 distribution by MLS method
3. A rule can be introduced which fits all the suggested distributions and finally suggests a single distribution suitable for all the seasons in the seasonal analysis of data
4. Further enhancement of the knowledge base of the ES
5. Enhancement of graphical capabilities like making use of microVAX-VMS graphics terminal
6. Implementation of the package in other computing environments and
7. It can be extended to the fitting of Time Series

REFERENCES

1 Abramowitz, M , and I A Stegun, 1965, " Handbook of Mathematical Functions ", Dover, New York, p 932

2 Ahmed, K et al, 1989, Analysis of Expert System for Fluid Flow in Networks Using a Qualitative Reasoning', Proc Hydrocomp' 89, Elscier pp 394-403

3 Arnold, C L , 1986,"Selection of Appropriate Water/Sanitation Treatment Technology in Developing Countries An Expert System Application", in 25 Not printed

4 Barnet, V , and T Lewis, 1983, "Outliers in Statistical Data', Wiley Series in Probability and Mathematical Statistics

5 Blom, G , 1958, ' Statistical Estimates and Transformed Beta Variables ", Wiley, New York, pp 68-75 and 143-146

6 Box, G E P and Cox, D R , 1964 "An Analysis of Transformation ", Journal of the Royal Statistical Society, Vol B26, pp 211-252

7 Chow, K C A , and W E Watt, 1990, " A Knowledge-Based Expert System for Flood Frequency Analysis ", Canadian Journal of Civil Engineering, Vol 17, pp 597-609

8 Chow, V T , 1964 " Handbook of Applied Hydrology", Mc-Graw Hill Book Co , New York

9 Chow V T , D R Maidment, L W Mayes, 1988 " Applied Hydrology", Mc-Graw Hill Publication, Singapore

10 Culbert C J , 1987 "CLIPS Reference Mannual" Version 4 0, NASA, Mission Support Directorate, Houston, Texas

11 Datta, B , and R C Peralta, "Expert System and Pattern Recognition to Identify Pollution Sources", World Water Issues in Evolution, ACSE, Vol 1, 1986

12 Fisher, R A , and L H C Tippett, 1928, " Limiting forms of the frequency distribution of the largest or smallest member of a sample ", Proc Cambridge Phil Soc , Vol 24, Part II, pp 180-191

13 Forgy, C L , 1984 "Rete A Fast Algorithm for the Many Pattern / Many Object Pattern Match Problem', Artificial Intelligence, in 16, pp 17-37

14 Foster, H A, 1924, " Theoretical frequency curves and their applications to engineering problems ", Trans. Am Soc. Civ. Engg., Vol 87, pp 142-173

15 Gasching, J et al, 1981 Development of a Knowledge Based System for Water Resources Problems", SRI Project 1619, SRI Intenational, August

16 Giarratano, J C , C Culbert, G Riley, and R Savely, 1987 "A Solution to the Expert System Delivery Problem ", NASA/Jhonson Space Centre, Artificial Intelligence Section FM7

17 Gringorten, I I , 1963, " A Plotting rule for extreme probability paper ", Jour Geophys Res , Vol 68, No 3, pp 813-814

18 Gupta, S C , and V K Kapoor, 1977 , "Fundamentals of Mathematical Statistics', New Delhi, Sultan Chand and Sons, India

19 Haan, C T , 1977 "Statistical Methods in Hydrology", The Iowa State University Press, Amer

20 Hawkins, D M , 1980, 'Identification of Outliers , London, New York, Chapman and Hall

21 Hayes-Roth, F , D A Waterman and D B Lenat, 1983 'Building Expert Systems ", Massachusetts, Addison-Wesley Publishers

22 Jackson, P , 1986 "Introduction to Expert System", Addison-Wesley Publishers

23 Jenkinson, A F , 1955, ' The frequency distribution of the annual maximum (or minimum) values of meteorological elements ", Quart Jour Roy Met Soc , Vol 81, pp. 158-171

24 Kite, G W , 1977 "Frequency and Risk Analysis in Hydrology , Water Resources Publications, Fort Collins, Colorado

25 Kostem, C N, and M L Maher, 1986 , " Expert Systems in Civil Engineering", Proc of Symposium Seattle, WA, USA

26 Krishnasami, M , 1976, ' Decoupled Multivariate Time Series Models for Multisite Stream flows", Ph D Thesis in Indian Institute of Technology, Kanpur, India

27 Levine, R I , D E Darang and B Edelson, 1988 " A Comprehensive Guide to AI and Expert System Turbo Pascal Edition", Mc Graw-Hill Book Company, Singapore

28 Maher M L , 1986 " Problem Solving Using Expert System Techniques "

29 Mishra, O.P , 1990, " An Approach to Frequency Analysis Using Expert System Technique ", M.Tech Thesis in Indian Institute of Technology, Kanpur, India

30 Oswald, O., 1989 " An Expert System for the Diagnosis of Tank Irrigation System", Ph D Thesis in Progress College of Engineering , Guindy, Madras, India

31 Ramaseshan, S , 1990 " Expert Systems for Issues in Water Resources Engineering ", Keynote address

32 Raudkivi, A J , 1979 " Hydrology ", Pergamon Press Ltd , Headington Hill Hall, Oxford, England

33 Shen, H W , 1976, " Stochastic Approaches to Water Resources ', Vol I , Fort Collins, Colorado

34 Yevjevich, V., 1972 " Probability and Statistics in Hydrology ", Water Resources Publication , Colorado, USA

APPENDIX A

```
                                        4-Dec-1991 10 11 46     VAX FORTRAN V4 6 244
                                       25-Nov-1991 14 24 14     $DISK3 [BADDU CLIPS JUNK]FORCLIP FOR,434

c       **********************************
        SUBROUTINE MLS(IOC,IOUTHI,IOUTLO,RXX,ILPT,Z,NC)
        COMMON/BLOCK1/NS,NPT,IALT
        COMMON/BLOCK2/RX(600),NX,HEADING(80)
        COMMON/GO/CSNS
        COMMON/PTG/GAMMA
        REAL KUT,KURT,M1,M2,M3
        DIMENSION X(60),XX(60),P(60),W(60),Y(60),Z(60),XL(60)
        DIMENSION TK(60),TPN1(60),TKLPN1(60)
        DIMENSION T(60),FX(60),XLP(60),XPL(60),SDP(60),SD(60)
        DIMENSION Z1(60)
        DIMENSION TPLP(60),PLN(60),J(20),X11(60),RESULT(20)
        DIMENSION IMAGE1(9000)
        DIMENSION SDP1(60),TX(60),EY1(80),EY3(80),Z3(60)
        DIMENSION X1(60),XL1(60),TKE1(60),PML(60),RXX(60),TX1(60)
        N=NX/NS
        LK=(NC-1)*N
        DO IJK=LK+1,LK+N
        X(IJK-LK)=RX(IJK)
        ENDDO
        NXX=N-1
        DO 6 K=1,NXX
        DO 6 I=K,NXX
        IF(X(K) GT X(I+1)) GO TO 6
        TEMP=X(K)
        X(K)=X(I+1)
        X(I+1)=TEMP
 6      CONTINUE
        GOTO (115,115,115,125,125,120,120)IOC
 115    WRITE(*,*)
        WRITE(*,*)           NORMAL AND LOG-NORMAL DISTRIBUTIONS'
        WRITE(*,*)'          ***********************************'
 99     AN1=FLOAT(N)+0 25
 750    DO 9 I=1,N
        IF(ILPT EQ 1)THEN
        P(I)=FLOAT(I)/FLOAT((N+1))
        ELSE
        P(I)=((FLOAT(I)-(3 0/8 0))/AN1)
        ENDIF
        IF(P(I) GT 0 5) GO TO 7
        W(I)=SQRT(ALOG(1 /(P(I)*P(I))))
        F=+1
        GO TO 8
 7      F=-1
        W(I)=SQRT(ALOG(1 /((1 -P(I))*(1 -P(I)))))
 8      Y(I)=W(I)-(2 515517+0.802853*W(I)+0 010328*W(I)*W(I))/
        1(1 +1 432788*W(I)+0 189269*W(I)*W(I)+0 001308*W(I)*W(I)*W(I))
        IF(ILPT EQ 1)THEN
        Z2(I)=F*Y(I)
        GOTO 9
        ENDIF
        Z(I)=F*Y(I)
 9      CONTINUE
        IF(ILPT EQ 1)GOTO 760
        IF(IOUT EQ 1)GOTO 939
        IF(ILPT EQ 2)RETURN
        IF(IOC EQ 3)THEN
```

```
              CALL LN3(1,M1,M2,M3,0,0)
              WRITE(A,A)' THE SHIFT PARAMETER IN LN3 IS ',M3
              DO I*1,N
              X(I)=X(I)-M3
              ENDDO
              ENDIF
              DO 4 I=1,N
              X1(I)=X(I)
              XL1(I)=ALOG(X(I))
       4      CONTINUE
              WRITE(A,10)
              WRITE(A,11)(I,X1(I),XL1(I),P(I),W(I),Z(I),I=IOUTHI+1,N-IOUTLO)
      10      FORMAT(3X, I',9X,'X(I)',9X, LX(I) ,9X, P(I) ,9X, W(I) ,9X, Z(I) /)
      11      FORMAT(1X,I3,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 3)
              NN=N-(IOUTHI+IOUTLO)
              DO I=IOUTHI+1,N-IOUTLO
              X1(I-IOUTHI)=X1(I)
              XL1(I-IOUTHI)=XL1(I)
              Z1(I-IOUTHI)=Z(I)
              ENDDO
              IF(IOC EQ 1)THEN
              CALL MLSREG(NN,Z1,X1,RESULT,IOC,1)
              DO I=1,N
              RXX(I)=RESULT(2)+Z(I)ARESULT(4)
              ENDDO
              ELSE
              CALL MLSREG(NN,Z1,XL1,RESULT,IOC,2)
              DO I=1,N
              RXX(I)=RESULT(2)+Z(I)ARESULT(4)
              ENDDO
              ENDIF
              IF(ILPT OF 3)RETURN
     760      DO I=1,N
              Z(I)=Z2(I)
              ENDDO
              RETURN
      C
     120      WRITE(A,A)
              WRITE(A,A)'             EXTREME-VALUE DISTRIBUTION'
              WRITE(A,A)'            AAAAAAAAAAAAAAAAAAAAAAAAAAAA
     770      DO 12 I=1,N
              AN1=FLOAT(N)+0 12
              IF(ILPT EQ 1)THEN
              P(I)=FLOAT(I)/FLOAT(N+1)
              ELSE
              PML(I)=((FLOAT(I)-0 44)/AN1)
              P(I)=PML(I)
              ENDIF
              T(I)=1 0/P(I)
              T1=ALOG(ALOG(T(I)/(T(I)-1 0)))
              IF(ILPT EQ 1)THEN
              TKE1(I)=-(0 5772+T1)ASQRT(6 0)/3 141562
              ELSE
              TK(I)=-(0 5772+T1)ASQRT(6 0)/3 141562
              ENDIF
              X1(I)=X(I)
              XL1(I)=ALOG(X(I))
```

```
          IF(ILPT EQ 1)XL1(I)=ALOG(X(I))
   12     CONTINUE
          IF(ILPT EQ 1)GOTO 780
          IF(ILPT EQ 2)THEN
          DO I=1,N
          Z(I)=TK(I)
          ENDDO
          RETURN
          ENDIF
          WRITE(A,13)
          WRITE(A,14)(I,X1(I),XL1(I),P(I),T(I),TK(I),I=IOUTHI+1,N-IOUTLO)
   13     FORMAT(3X,'I',9X, X(I) ,9X,'LX(I)',9X, P(I) ,9X,'T(I)',9X, KT(I)'/)
   14     FORMAT(1X,I3,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 2)
          NN=N-(IOUTHI+IOUTLO)
          DO I=IOUTHI+1,N-IOUTLO
          X1(I-IOUTHI)=X1(I)
          XL1(I-IOUTHI)=XL1(I)
          PML(I-IOUTHI)=PML(I)
          ENDDO
          IF(IOC EQ 6)THEN
          CALL PARAM(X1,NN,AME,VAR,SKW,KUT,STD,2)
          E1SUM=0 0
          B1=STD*SQRT(6 0)/3 1415
          A1=AME-0 5772*B1
          DO I=1,NN
          EY1(I)=0 0
          EY1(I)=EY1(I)+EXP(-(X1(I)/B1))
          E1SUM E1SUM+EY1(I)
          ENDDO
          NIL=0
   62     NXX=NN-1
          DO 61 K=1,NXX
          DO 61 I=K,NXX
          IF(EY1(K) GT EY1(I+1)) GO TO 61
          TEMP=EY1(K)
          EY1(K)=EY1(I+1)
          EY1(I+1)=TEMP
   61     CONTINUE
          IF(NIL EQ 1)GOTO 63
          P1=0 0
          DO I=1,NN
          P(I)=ALOG(PML(I))
          P1=P1+P(I)
          ENDDO
          ALAM=E1SUM/P1
          WRITE(A,A)' FOR THE EXPONENTIAL DISTRIBUTION '
          WRITE(A,A)'THE METHOD OF LEAST SQUARE ERRORS CONSTANT = ',ALAM
          DO I=1,N
          EY1(I)=P(I)*ALAM
          EY1(I)=ALOG(EY1(I))*(-B1)
          ENDDO
          NIL=1
          GOTO 62
   63     DO I=1,N
          RXX(I)=EY1(I)
          ENDDO
          CALL PARAM(EY1,N,AMEAN,VAR,SKEW,KURT,STDEV,2)
```

```
        CALL CHISORT(8,AMEAN,STDEV,EY1,N,EY1,NS,NPT,IALT,IOC,1)
        WRITE(A,A)' THE CHI-SQUARE VALUE OBTAINED USING MLS DUE TO'
        WRITE(A,A)'          THE PRESENCE OF OUTLIERS IS   '
        CONTINUE
        ELSE
        CALL PARAM(XL1,NN,AMEAN,VARI,SKEW,KURT,STDEV,2)
        E3SUM=0 0
        B1=STDEVASQRT(6 0)/3 1415
        A1=AMEAN-0 5772AB1
        DO I=1,NN
        EY3(I)=0 0
        EY3(I)=EY3(I)+EXP(-(XL1(I)/B1))
        E3SUM=E3SUM+EY3(I)
        ENDDO
        NIL1=0
64      NXX=NN-1
        DO 60 K=1,NXX
        DO 60 I=K,NXX
        IF(EY3(K) GT EY3(I+1)) GO TO 60
        TEMP=EY3(K)
        EY3(K)=EY3(I+1)
        EY3(I+1)=TEMP
60      CONTINUE
        IF(NIL1 EQ 1)GOTO 65
        P1=0 0
        DO I=1,NN
        P(I)=ALOG(PML(I))
        P1=P1+P(I)
        ENDDO
        ALAM=E3SUM/P1
        WRITE(A,A)' FOR THE EXPONENTIAL DISTRIBUTION '
        WRITE(A,A)'THE METHOD OF LEAST SQUARE ERRORS CONSTANT = ',ALAM
        DO I=1,N
        EY3(I)=P(I)AALAM
        EY3(I)=ALOG(ABS(EY3(I)))A(-B1)
        ENDDO
        NIL1=1
        GOTO 64
65      DO I=1,N
        RXX(I)=EY3(I)
        ENDDO
        CALL PARAM(EY3,N,AMEAN,VAR,SKEW,KURT,STDEV,2)
        CALL CHISORT(8,AMEAN,STDEV,EY3,N,EY3,NS,NPT,IALT,IOC,1)
        WRITE(A,A)' THE CHI-SQUARE VALUE OBTAINED USING MLS DUE TO'
        WRITE(A,A)'          THE PRESENCE OF OUTLIERS IS   '
        CONTINUE
        ENDIF
        IF(ILPT.EQ 3)RETURN
780     DO I=1,N
        Z(I)=TKE1(I)
        ENDDO
        RETURN
C
125     WRITE(A,A)
        WRITE(A,A)'        PEARSON AND LOG-PEARSON DISTRIBUTION'
        WRITE(A,A)'       AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA'
        DO I=1,N
```

```
        TX1(I)=X(I)
        ENDDO
126     DO I=1,N
        XPL(I)=ALOG(X(I))
        XL(I)=ALOG10(X(I))
        ENDDO
        CALL PARAM(X,N,PMEAN,PVAR,CSX,PKURT,SDX,4)
        CALL PARAM(XL,N,PLMEAN,PLVAR,CSPL,PLKURT,SDPL,4)
        DO 19 I=1,N
        AN1=FLOAT(N)+0 25
        P(I)=((FLOAT(I)-(3 0/8 0))/AN1)
19      CONTINUE
        PLK=CSPL/6 0
        PK=CSX/6 0
        IOU1=1
        GOTO 99
929     DO 21 I=1,N
        TK1=Z(I)+(Z(I)**2-1 0)*PLK+1 0/3 0*(Z(I)**3-6 0*Z(I))*PLK**2
        TK2=-(Z(I)**2-1 0)*PLK**3+Z(I)*PLK**4+1 0/3 0*PLK**5
        TK3=Z(I)+(Z(I)**2-1 0)*PK+1 0/3 0*(Z(I)**3-6 0*Z(I))*PK**2
        TK4=-(Z(I)**2-1 0)*PK**3+Z(I)*PK**4+1 0/3 0*PK**5
        TKLP(I)=TK1+TK2
        TK(I)=TK3+TK4
21      CONTINUE
        IOU1=2
        IF(ILPT EQ 2)THEN
        IF(IOC EQ 4)THEN
        DO I=1,N
        Z(I)=TK(I)
        ENDDO
        RETURN
        ELSE
        DO I=1,N
        Z(I)=TKLP(I)
        ENDDO
        RETURN
        ENDIF
        ENDIF
        WRITE(A,22)
        WRITE(A,23)(I,X(I),XL(I),P(I),TK(I),TKLP(I),I=IOUTHI+1,N-IOUTLO)
22      FORMAT(3X,'I',9X,'X(I)',9X,'LX(I)',9X,'P(I)',7X,'KTP(I)',
        17X,'KTLP(I)'/)
23      FORMAT(1X,I3,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 2,5X,F8 2)
        NN=N-(IOUTHI+IOUTLO)
        DO I=IOUTHI+1,N-IOUTLO
        X1(I-IOUTHI)=X(I)
        XL1(I-IOUTHI)=XL(I)
        TKN1(I-IOUTHI)=TK(I)
        TKLPN1(I-IOUTHI)=TKLP(I)
        ENDDO
        IF(IOC EQ 4)THEN
        CALL HLSREG(NN,TKN1,X1,RESULT,IOC,1)
        DO I=1,N
        RXX(I)=RESULT(2)+TK(I)*RESULT(4)
        ENDDO
        ELSE
        CALL HLSREG(NN,TKLPN1,XL1,RESULT,IOC,2)
```

```
      DO I=1,N
      RXX(I)=RESULT(2)+TKLP(I)*RESULT(4)
      ENDDO
      ENDIF
      IF(CSNS EQ 1)GOTO 876
      IF(IDC EQ 4)THEN
      WRITE(*,*)' PEARSON TYPE 3 DISTRIBUTION '
      WRITE(*,*)' *************************** '
      BETA=(2 0/CSX)**2
      ESK=CSX
      ELSE
      WRITE(*,*)' LOG-PEARSON TYPE 3 DISTRIBUTION '
      WRITE(*,*)  ******************************* '
      BETA=(2 0/CSPL)**2
      ESK=CSPL
      ENDIF
      ALPHA=RESULT(4)/SQRT(BETA)
      GAMMA=RESULT(2)-RESULT(4)*SQRT(BETA)
      WRITE(*,*)
      WRITE(*,*)'  METHOD OF LEAST SQUARES '
      WRITE(*,*)'  ----------------------
      WRITE(*,*)
      WRITE(*,*)' -------------------------------------- '
      WRITE(*,*)
      WRITE(*,*)' ALPHA = ',ALPHA,'    MEAN  = ',RESULT(2)
      WRITE(*,*)' BETA  = ',BETA,'     STDEV = ',RESULT(4)
      WRITE(*,*)' GAMMA = ',GAMMA,     SKEW  = ',ESK
      WRITE(*,*)
      WRITE(*,*)' __  _____ _____ ___ _  __ ___________       '
876   IF(ILPT EQ 3)RETURN
      END
```

```
C       ****************************************************
        SUBROUTINE FREQPL(N,X)
C       THIS ROUTINE DRAWS A HISTOGRAM FOR THE GIVEN DATA
C       ****************************************************
        DIMENSION X(50),X1(50),J(10)
        CHARACTER STRA(10)
        INCLUDE '($SMGDEF)'
        INTEGER SMG$CREATE_VIRTUAL_DISPLAY,SMG$CREATE PASTEBOARD
        INTEGER SMG$DRAW_LINE,SMG$PUT CHARS
        INTEGER SMG$PASTE VIRTUAL DISPLAY,SMG$DRAW_RECTANGLE
        INTEGER DISPLAY1,PASTE1,ROWS,COLUMNS,STATUS
        STATUS=SMG$DELETE_PASTEBOARD(PASTE1)
        NX=N-1
        DO 6 K=1,NX
        DO 6 I=K,NX
        IF(X(K) GT X(I+1)) GO TO 6
        TEMP-X(K)
        X(K)=X(I+1)
        X(I+1)=TEMP
 6      CONTINUE
        DO 1 I=1,N
        X1(I)=X(I)/1000 0
 1      CONTINUE
C
        DX1=(X(1)-X(N))/700 0
        DX=(FLOAT(INT(DX1)))/10 0
        Y1=FLOAT(INT(X1(N)))
        IF(INT(Y1/10 0) EQ 0)THEN
        IY1=INT(Y1*10 0)
        IDX=INT(20*DX)
        ELSE
        IY1=INT(Y1)
        IDX=INT(2*DX)
        ENDIF
C
        LC=0
        DO 15 I=1,10
        Y2=Y1+DX
        J(I)=0
        DO 20 K=1,N
        IF(X1(K) GT Y1 AND X1(K) LT Y2)J(I)=J(I)+1
 20     CONTINUE
        LC=LC+1
        IF(Y2 GT X1(1))GOTO 2
        Y1=Y2
 15     CONTINUE
C
C
 2      ROWS=25
        COLUMNS=65
        STATUS=SMG$CREATE_VIRTUAL DISPLAY
     1                    (ROWS,COLUMNS,DISPLAY1,SMG$M_BORDER)
        IF( NOT STATUS) CALL LIB$SIGNAL(%VAL(STATUS))
C
        STATUS=SMG$CREATE_PASTEBOARD (PASTE1)
        IF( NOT.STATUS) CALL LIB$SIGNAL(%VAL(STATUS))
C
```

```
      STATUS=SMG$DRAW LINE (DISPLAY1,18,7,18,60)
      STATUS=SMG$DRAW_LINE (DISPLAY1,1,7,18,7)
      LC1=7
      LC2=12
      DO 30 I=1,LC
      LR1=18-J(I)
      IF(J(I) GT 18)LR1=18
      IF(LR1 EQ 18)THEN
      STATUS=SMG$DRAW_LINE (DISPLAY1,18,LC1,18,LC2)
      GOTO 31
      ENDIF
      STATUS=SMG$DRAW_RECTANGLE (DISPLAY1,LR1,LC1,18,LC2)
31    LC1=LC2
      LC2=LC2+5
30    CONTINUE
      IF( NOT STATUS) CALL LIB$SIGNAL(%VAL(STATUS))
C
      STATUS=SMG$PUT CHARS (DISPLAY1,
     1                'HISTOGRAM FOR THE GIVEN DATA ,21,20)
C
      DO 61 K=1,8
      DO 61 I=K,8
      IF(J(K) GT J(I+1)) GO TO 61
      TEMP=J(K)
      J(K)=J(I+1)
      J(I+1)=TEMP
 61   CONTINUE
C
      DO 40 I=1,J(1)+2
      KR=18-I
      K=I
      STATUS=SMG$DRAW_LINE (DISPLAY1,KR,6,KR,7)
40    CONTINUE
C
      DO 130 I=1,J(1)+2
      IF (I/10  EQ  0) THEN
      STRA(1)=CHAR(32)
      ELSE
      STRA(1) = CHAR(I/10+48)
      ENDIF
      STRA(2) = CHAR(I-I/10*10+48)
      STATUS=SMG$PUT_CHARS (DISPLAY1,STRA,18-I,4)
 130  CONTINUE
C
      DO 131 I=1,7
      STRA(2) = CHAR(IY1-IY1/10*10+48)
      STRA(1) = CHAR(IY1/10+48)
      STATUS=SMG$PUT_CHARS (DISPLAY1,STRA,19,(I-1)*10+7)
      IY1 = IY1+IDX
 131  CONTINUE
C
      STATUS=SMG$PUT_CHARS (DISPLAY1,'F',9,1)
      STATUS=SMG$PUT CHARS (DISPLAY1,'R',10,1)
      STATUS=SMG$PUT CHARS (DISPLAY1,'E',11,1)
      STATUS=SMG$PUT CHARS (DISPLAY1,'Q ,12,1)
      STATUS=SMG$PUT CHARS (DISPLAY1,'U',13,1)
      STATUS=SMG$PUT_CHARS (DISPLAY1,'E ,14,1)
```

```
          STATUS=SMG$PUT CHARS (DISPLAY1, H',15,1)
          STATUS=SMG$PUT_CHARS (DISPLAY1, C ,16,1)
          STATUS=SMG$PUT_CHARS (DISPLAY1,'Y',17,1)
C
          STATUS=SMG$PASTE_VIRTUAL_DISPLAY (DISPLAY1,PASTE1,4,15)
          IF( NOT STATUS) CALL LIB$SIGNAL(%VAL(STATUS))
          WRITE(A,100)
100       FORMAT(1H1)
          RETURN
3         END
```

```
C       AAAAAAAAAAAAAAAAAAAAAAAAAAA
C       CHECKS FOR OUTLIERS IN ANY DISTRIBUTION
        SUBROUTINE OUTLIER(IOUTHI,IOUTLO,NC)
        COMMON/BLOCK1/NS,NPT,IALT
        COMMON/BLOCK2/RX(600),NX,HEADING(80)
        REAL KURT,YO,KUT,M1,M2,M3
        DIMENSION X(60),TX(60),TITLE(80),XO(101),EY1(60),EY3(60)
        DIMENSION TEST(100),SPH(59),SPL(18)
        DATA  YO  / 2 036,2 088,2 134,2 175,2 213,2 247,2 279,2 309,2 335
        1,2 361,2 385,2 408,2 429,2 448,2 467,2 486,2 502,2 519,2 534
        2,2 549,2 563,2.577,2 591,2 604,2 616,2 628,2 639,2 650,2 661
        3,2 671,2 682,2.693,2 7,2 71,2 719,2 727,2 736,2 744,2 753,2 760
        4,2 768,2 775,2 782,2 789,2 796,2 804,2 81,2 817,2 8238,2 83
        5,2 837,2 842,2 848,2 854,2 86,2 866,2 871,2 876,2 882,2 887
        6,2 893,2 898,2 9,2 907,2 912,2 917,2 921,2 926,2 93,2 935
        7,2 94,2 944,2 948,2 952,2 956,2 957,2 96,2 965,2 969,2 973
        8,2 977,2 981,2 985,2 989,2 99,2 996,3 0,3 0,3 0,3 01,3 013
        9,3 017,3 02,3 023,3 026,3 029,3 033,3 036,3 039,3.042,3 045/
        DATA  SPH  / 0 9750,0 8709,0 7679,0 6838,0 6161,0 5612,0 5157
        1,0 4775,0 4450,0 4187,0 3924,0 37313,0 3538,0 3346,0 32178
        2,0 30896,0 29614,0 28332,0 2705,0 26173,0 25295,0 24417,0 2354
        3,0 22916,0 22293,0 2167,0 21046,0 20423,0 1980,0 19396,0 18992
        4,0 18588,0 18184,0 1778,0 17376,0 16972,0 16568,0 16164,0 1576
        5,0 15537,0 15315,0 15093,0 1487,0 14647,0 14425,0 14203,0 1398
        6,0 13757 0 13535,0 13313,0 1309,0 12867,0 12645,0 12423,0 1220
        7,0 11977,0 11755,0 11533,0 1131/
        DATA  SPL  / 0 00844,0 00424,0 00255,0 00170,0 00122,0 000913
        1,0 00071,0 000568,0 000478,0 000388,0 0003345,0 000281
        2,0 000247,0 000213,0 000190,0 000167,0 000151,0.000135/
        N=NX/NS
        LK=(NC-1)*N
        DO IJK=LK+1,LK+N
        X(IJK-LK)=RX(IJK)
        ENDDO
        NXX=N-1
        DO 6 K=1,NXX
        DO 6 I=K,NXX
        IF(X(K) GT X(I+1)) GO TO 6
        TEMP=X(K)
        X(K)=X(I+1)
        X(I+1)=TEMP
   6    CONTINUE
        SS=0 0
        DO I=1,N
        SS=SS+X(I)
        ENDDO
        CALL PARAM(X,N,AME,VARI,SKW,KUT,STD,2)
        REWIND 20
        WRITE(*,*)
        WRITE(*,*)  THE DATA IS BEING CHECKED FOR OUTLIERS'
        WRITE(*,*)' FOR WHICH DISTRIBUTION DO YOU WANT IT CHECKED i e '
        WRITE(*,*)' WHETHER 1  NORMAL,  2  LOG-NORMAL(2 PARAMETERS) '
        WRITE(*,*)' 3  LOG-NORMAL(3 PARAMETERS),  4. EV-I '
        WRITE(*,*)' 5  EV-III,  6  PT-III,  7  LPT-III '
        READ(*,*)IOU
        GOTO (110,15,15,115,15,130,15)IOU
  15    IF(IOU EQ 3)THEN
```

```
        CALL LN3(1,M1,M2,M3,0,0)
        X(I)=X(I)-M3
        ENDIF
        DO 212 I=1,N
        TX(I)=X(I)
        IF(TX(I) LE 0 0) TX(I)=0 1
212     X(I)=ALOG(TX(I))
        SS=0 0
        DO I=1,N
        SS=SS+X(I)
        ENDDO
        CALL PARAM(X,N,AMEAN,VAR,SKEW,KURT,STDEV,2)
        IF(IOU EQ 2 OR IOU EQ 3)GOTO 142
        IF(IOU EQ 5)GOTO 120
        IF(IOU EQ 7)GOTO 131
110     WRITE(A,A)
        WRITE(A,A)  THE PARAMETERS FOR THE ORIGINAL DATA ARE  '
        WRITE(A,192)AME,SKW,STD
192     FORMAT(2X, MEAN=',2X,E10 4,/,2X, SKEW=',2X,E12 5,/,
        12X, STDEV=',2X,E12 5)
142     TYPE A,' AAA FOR ORIGINAL DATA AAA'
        TYPE A,'CHECK FOR OUTLIERS BY WRC METHOD'
C       THE RANGE OF A DATA
        XMIN=1000000000 0
        XMAX=0 00000001
        IOUTHI=0
        IOUTLO=0
        NN=N-9
        DO I=1,N
        IF(X(I) LT XMIN)XMIN=X(I)
        ENDDO
        DO I=1,N
        IF(X(I) GT XMAX)XMAX=X(I)
        ENDDO
        WRITE(A,A)
        WRITE(A,A)' THE PARAMETERS FOR THE LOG-TRANSFORMED DATA ARE  '
        WRITE(A,190)AMEAN,SKEW,STDEV
190     FORMAT(2X,'MEAN=',2X,E10 4,/,2X,'SKEW=',2X,E12 5,/,2X,
        1'STDEV=',2X,E12 5,/)
        WRITE(A,200)XMIN,XMAX
200     FORMAT(2X,'XMIN= ',E12 5,5X,'XMAX= ',E12 5)
        IF(IOU EQ 1 OR IOU EQ 6 OR IOU EQ 7)THEN
        AM=AME
        ST=STD
        ELSE
        AM=AMEAN
        ST=STDEV
        ENDIF
        YH=AM+KO(NN)AST
        YL=AM-KO(NN)AST
        IF (IOU EQ 2 OR IOU EQ 3)THEN
        YH1=EXP(YH)
        YL1=EXP(YL)
        WRITE(A,A)' FOR LOG-NORMAL DISTRIBUTION
        WRITE(A,A)' THE THRESHOLD VALUE FOR HIGHER OUTLIERS    ,YH1
        WRITE(A,A)' THE THRESHOLD VALUE FOR LOWER OUTLIERS    ,YL1
        XMIN=EXP(XMIN)
```

```
          XMAX=EXP(XMAX)
          ELSE
          WRITE(*,*)' FOR NORMAL DISTRIBUTION '
          WRITE(*,*)  THE THRESHOLD VALUE FOR HIGHER OUTLIERS   ',YH
          WRITE(*,*)  THE THRESHOLD VALUE FOR LOWER OUTLIERS   ',YL
          ENDIF
          DO 3 I=1,N
          IF(X(I) GE YH)IOUTHI=IOUTHI+1
          IF(X(I) LE YL)IOUTLO=IOUTLO+1
    3     CONTINUE
          WRITE(*,100)XMIN,XMAX,IOUTHI,IOUTLO
  100     FORMAT(1X,'XMIN= ,E12 5,2X,'XMAX= ,E12 5,/,1X,
          1'NO OF HIGH OUTLIERS =  ,I5,/,1X, NO OF LOW OUTLIERS= ',I5)
          IF(IOU EQ 6 OR IOU EQ 7)GOTO 501
          IF(IOU EQ 2 OR IOU EQ 3)THEN
          DO I=1,N
          IF(IOU EQ 3)X(I)=X(I)+M3
          X(I)=EXP(X(I))
          ENDDO
          ENDIF
          IF(IOU EQ 1)THEN
          CALL PAEST(IOU,N,X,IOUTHI,IOUTLO,1)
          ELSE
          CALL PAEST(IOU,N,X,IOUTHI,IOUTLO,I)
          ENDIF
          GOTO 500
C         DATA IS INITIALLY TRANSFORMED FROM EV-I TYPE TO EXPONENTIAL TYPE
C         AND THEN TESTED FOR OUTLIERS
  115     E1SUM=0 0
          B1=STDASORT(6 0)/3 1415
          A1=AME-0 5772AB1
          WRITE(*,*) A=  ,A1,' B= ',B1
          DO I=1,N
          EY1(I)=0 0
          EY1(I)=EY1(I)+EXP(-(X(I)/B1))
          E1SUM=E1SUM+EY1(I)
          ENDDO
          WRITE(*,*)' E1SUM =  ,E1SUM
          NXX=N-1
          DO 61 K=1,NXX
          DO 61 I=K,NXX
          IF(EY1(K) GT EY1(I+1)) GO TO 61
          TEMP=EY1(K)
          EY1(K)=EY1(I+1)
          EY1(I+1)=TEMP
   61     CONTINUE
          ALAMB1=EXP(-(A1/B1))
          ALAMB2=EXP(A1/B1)
          WRITE(*,*)' FOR THE EXPONENTIAL DISTRIBUTION '
          WRITE(*,*)' ORIGIN = 0 '
          WRITE(*,*)' SCALE PARAMETER FOR GREATEST VALUE DIST  = ',ALAMB1
          WRITE(*,*)' SCALE PARAMETER FOR SMALLEST VALUE DIST  = ',ALAMB2
          GOTO 510
  120     E3SUM=0 0
          B3=STDEVASORT(6 0)/3 1415
          A3-AMEAN-0 5772AB3
          WRITE(*,*)'A= ',A3,' B= ',B3
```

```
            DO I=1,N
            EY3(I)=0 0
            EY3(I)=EY3(I)+EXP(-(X(I)/B3))
            E3SUM=E3SUM+EY3(I)
            ENDDO
            WRITE(A,A)' E3SUM =  ,E3SUM
            NXX=N-1
            DO 62 K=1,NXX
            DO 62 I=K,NXX
            IF(EY3(K) GT EY3(I+1)) GO TO 62
            TEMP=EY3(K)
            EY3(K)=EY3(I+1)
            EY3(I+1)=TEMP
      62    CONTINUE
            ALAMB3=EXP(-A3/B3)
            ALAMB4-EXP(A3/B3)
            WRITE(A,A)  FOR THE EXPONENTIAL DISTRIBUTION
            WRITE(A,A)' ORIGIN = 0 '
            WRITE(A,A)  SCALE PARAMETER FOR GREATEST VALUE DIST  = ',ALAMB3
            WRITE(A,A)' SCALE PARAMETER FOR SMALLEST VALUE DIST  =  ,ALAMB4
            GOTO 510
      C     TEST FOR OUTLIERS IN THE EXPONENTIAL DISTRIBUTION
      C     CONCECUTIVE TESTS FOR A SINGLE UPPER OUTLIER IN EXPONENTIAL DIST
      C     AN UPPER OUTLIER IN THE EXT  VALUE DIST  IS TRANSFORMED AS
      C     LOWER OUTLIER IN THE EXPONENTIAL DISTRIBUTION
      510   N1=N/2
            N11=N
            N2=N
            IOUTHI=0
            IOUTLO=0
            IF(IOU EQ 4)THEN
            SSH=E1SUM
            DO I=1,N1
            TEST(I)=EY1(I)/SSH
            IF(N11 GT 120)THEN
            SPH1=N11*1 0/((1 0+(TEST(I)/(1 0-TEST(I)))))**(N11-1)
            ENDIF
            IF(N11 GT 60 AND N11 LE 120)THEN
            SPH1=0 1131-((0 1131-0 0632)/60 0)*(N11-60)
            ENDIF
            N12=N11-1
            IF(N12 GT 59)THEN
            SPH(N12)=0 0
            ELSE
            SPH1=0 0
            ENDIF
            IF(SPH(N12) LT.TEST(I) AND SPH1 EQ 0 OR
           1SPH(N12) EQ 0 AND SPH1 LT TEST(I))THEN
            IOUTLO IOUTLO+1
            SSH=SSH-EY1(I)
            N11=N11-1
            GOTO 151
            ELSE
            GOTO 159
            ENDIF
      151   ENDDO
            ELSE
```

```
              SSH=E3SUM
              DO I=1,N1
              TEST(I)=EY3(I)/SSH
              IF(N11 GT 120)THEN
              SPH1=N11A1 0/((1 0+(TEST(I)/(1 0-TEST(I)))))AA(N11-1)
              ENDIF
              IF(N11 GT 60 AND N11 LE 120)THEN
              SPH1=0 1371-((0 1371-0 0759)/60 0)A(N11-50)
              ENDIF
              N12=N11-1
              IF(N12 GT 59)THEN
              SPH(N12)=0 0
              ELSE
              SPH1=0 0
              ENDIF
              IF(SPH(N12) LT TEST(I) AND SPH1 EQ 0 OR
              1SPH(N12) EQ 0 AND SPH1 LT.TEST(I))THEN
              IOUTLO=IOUTLO+1
              SSH=SSH-EY3(I)
              N11=N11-1
              GOTO 153
              ELSE
              GOTO 159
              ENDIF
      153     ENDDO
              ENDIF
      159     WRITE(A,A)' NO  OF LOW OUTLIERS IN THE EV DIST =  ',IOUTLO
              WRITE(A,A)
              GOTO 520
      C       CONCECUTIVE TESTS FOR A SINGLE LOWER OUTLIER IN EXPONENTIAL DIST
      C       AN LOWER OUTLIER IN THE EXT  VALUE DIST  IS TRANSFORMED AS
      C       UPPER OUTLIER IN THE EXPONENTIAL DISTRIBUTION
      520     N11=N
              IF(IOU EQ 4)THEN
              SS=E1SUM
              DO I=N2,N1,-1
              TEST(I)=EY1(I)/SS
              IF(N11 GT 100)THEN
              SPL1=N11A1 0/((1 0+(TEST(I)/(1 0-TEST(I)))))AA(N11-1)
              ENDIF
              IF(N11 GT 20 AND N11 LE 30)THEN
              SPL1=0 000135-((0 000135-0 0000589)/10)A(N11-20)
              ENDIF
              IF(N11 GT 30 AND N11 LE 40)THEN
              SPL1=0 0000589-((0 0000589-0 0000329)/10)A(N11-30)
              ENDIF
              IF(N11 GT 40 AND N11 LE 50)THEN
              SPL1=0 0000329-((0 0000329-0 0000209)/10)A(N11-40)
              ENDIF
              IF(N11 GT 50 AND N11 LE 100)THEN
              SPL1=0 0000209-((0 0000209-0 0000518)/50)A(N11-50)
              ENDIF
              N12=N11-2
              IF(N12 GT 18)THEN
              SPL(N12)=0 0
              ELSE
              SPL1=0 0
```

```
          ENDIF
          IF(SPL(N12) GT TEST(I) AND SPL1 EQ 0 OR
         1SPL(N12) EQ 0 AND SPL1 GT TEST(I))THEN
          IOUTHI=IOUTHI+1
          SS=SS-EY1(I)
          IF(I EQ 1)GOTO 158
          N11=N11-1
          GOTO 152
          ELSE
          GOTO 158
          ENDIF
152       ENDDO
          ELSE
          SS=E3SUM
          DO I=N2,N1,-1
          TEST(I)=EY3(I)/SS
          IF(N11 GT 100)THEN
          SPL1=N11*1 0/((1 0+(TEST(I)/(1 0-TEST(I)))))**(N11-1)
          ENDIF
          IF(N11 GT 20 AND N11 LE.30)THEN
          SPL1=0 0000264-((0 0000264-0 0000116)/10)*(N11-20)
          ENDIF
          IF(N11 GT 30 AND N11 LE.40)THEN
          SPL1=0 0000116-((0 0000116-0.00000644)/10)*(N11-30)
          ENDIF
          IF(N11 GT 40 AND N11 LE 50)THEN
          SPL1=0 00000644-((0 00000644-0 00000410)/10)*(N11-40)
          ENDIF
          IF(N11 GT 50.AND N11 LE 100)THEN
          SPL1=0 0000041-((0 0000041-0 00000102)/50)*(N11-50)
          ENDIF
          N12=N11-2
          IF(N12 GT 18)THEN
          SPL(N12)=0 0
          ELSE
          SPL1=0 0
          ENDIF
          IF(SPL(N12) GT TEST(I) AND SPL1 EQ 0.OR
         1SPL(N12) EQ 0 AND SPL1 GT TEST(I))THEN
          IOUTHI=IOUTHI+1
          SS=SS-EY3(I)
          IF(I EQ 1)GOTO 158
          N11=N11-1
          GOTO 154
          ELSE
          GOTO 158
          ENDIF
154       ENDDO
          ENDIF
158       WRITE(*,*)' NO  OF HIGH OUTLIERS IN THE EV DIST =  ',IOUTHI
          IF(IOU EQ 4)THEN
          CALL PAEST(IOU,N,X,IOUTHI,IOUTLO,1)
          ELSE
          DO I=1,N
          X(I)=EXP(X(I))
          ENDDO
          CALL PAEST(IOU,N,X,IOUTHI,IOUTLO,1)
```

```
              ENDIF
              GOTO 500
      C       TRANSFORM P-III DIST TO NORMAL DIST AND THEN CHECK FOR OUTLIERS
      130     CALL PARAM(X,N,AME,VARI,SKW,KUT,STD,2)
              WRITE(6,*)
              WRITE(6,*)' THE PARAMETERS FOR THE ORIGINAL DATA ARE  '
              WRITE(6,191)AME,SKW,STD
      191     FORMAT(2X,'MEAN=',2X,F10 4,/,2X,'SKEW=',2X,E12 5,/,
              12X, STDEV=',2X,E12 5)
              WRITE(6,*)
      131     DO I=1,N
              TX(I)=X(I)
              ENDDO
              IF(IOU EQ 6)THEN
              CALL IPT(N,1,AME,STD,SKW,X)
              CALL PARAM(X,N,AME,VARI,SKW,KUT,STD,2)
              GOTO 142
              ELSE
              CALL IPT(N,1,AMEAN,STDEV,SKEW,X)
              CALL PARAM(X,N,AME,VARI,SKW,KUT,STD,2)
              GOTO 142
              ENDIF
      501     IF(IOU EQ 7)THEN
              DO I=1,N
              TX(I)=EXP(TX(I))
              ENDDO
              ENDIF
              IF(IOU EQ 6)THEN
              CALL PAEST(IOU,N,TX,IOUTHI,IOUTLO,1)
              ELSE
              CALL PAEST(IOU,N,TX,IOUTHI,IOUTLO,1)
              ENDIF
      500     RETURN
              END
```

CE-1991-M-KAL-EXP